



Transmission and Storage Rates for Sequential Massive Random Access

Elsa Dupraz, Thomas Maugey, Aline Roumy, Michel Kieffer

► To cite this version:

Elsa Dupraz, Thomas Maugey, Aline Roumy, Michel Kieffer. Transmission and Storage Rates for Sequential Massive Random Access. 2018. hal-01799016

HAL Id: hal-01799016

<https://inria.hal.science/hal-01799016>

Preprint submitted on 24 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Transmission and Storage Rates for Sequential Massive Random Access

Elsa Dupraz¹, Thomas Maugey², Aline Roumy², Michel Kieffer³

¹ Telecom Bretagne; UMR CNRS 6285 Lab-STICC, ² INRIA Rennes Bretagne-Atlantique, ³ L2S, UMR CNRS 8506; CentraleSupélec; Univ. Paris-Sud

Abstract

This paper introduces a new source coding paradigm called Sequential Massive Random Access (SMRA). In SMRA, a set of correlated sources is encoded once for all and stored on a server, and clients want to successively access to only a subset of the sources. Since the number of simultaneous clients can be huge, the server is only allowed to extract a bitstream from the stored data: no re-encoding can be performed before the transmission of the specific client's request. In this paper, we formally define the SMRA framework and introduce both storage and transmission rates to characterize the performance of SMRA. We derive achievable transmission and storage rates for lossless source coding of i.i.d. and non i.i.d. sources, and transmission and storage rates-distortion regions for Gaussian sources. We also show two practical implementations of SMRA systems based on rate-compatible LDPC codes. Both theoretical and experimental results demonstrate that SMRA systems can reach the same transmission rates as in traditional point to point source coding schemes, while having a reasonable overhead in terms of storage rate. These results constitute a breakthrough for many recent data transmission applications in which different parts of the data are requested by the clients.

Index Terms

Massive Random Access, Free-viewpoint Television, Video Compression, Sensor Networks, Lossless and Lossy Source Coding.

I. INTRODUCTION

According to [1], the amount of data available on the web is growing exponentially. In the huge databases of, *e.g.*, videos or pictures stored in online servers, the data usually contains redundancies. Efficient coding algorithms have been developed since decades to exploit these redundancies in order to limit the tremendous increase of storage needs. In applications such as free-viewpoint television (FTV) or when exploiting data collected by a network of sensors, the items of the database (the different views of a video or the signals captured by different sensors) show strong correlations with each other. In such applications, all the database should be compressed jointly in order to exploit all the redundancies and hence to maximize the storage efficiency.

In emerging applications such as FTV, the data is massively and interactively accessed by heterogeneous clients. The explosion of the data volume makes unrealistic and overall not desirable for the client the reception of the whole dataset. For example, in FTV [2] or in sensor networks, the client is only interested by a subset of the views or of the collected data, and this subset may differ among clients. In such schemes, a large number of sources are jointly compressed, and different subsets of them must be transmitted to the clients, depending on their needs. In order for traditional coding schemes to be efficient both in terms of storage and transmission rates, the server would need to decode the whole stream and re-encode only the requested sources. However, in this context, re-encoding is not desirable due to the potentially large number of simultaneous clients' requests. As a result, for this new context of data delivery, traditional coding schemes cannot minimize both the storage rate and the transmission rate without re-encoding. Hence, there is a need for the development of new performing coding algorithms that take into account the variety of client's requests.

While some applications of the described scenario already exist, they have never been formally studied, and more particularly their transmission and storage performance have never been derived. In this paper, we propose a formal definition and a theoretical analysis of this new data coding and delivery scheme where only part of the sources are requested by the clients. We call this paradigm, the *Sequential Massive Random Access (SMRA)*, where "Random" refers to the uncertainty in the clients requests, "Massive" describes the fact that the number of clients requesting data is high and makes the data re-encoding unfeasible, and "Sequential" means that clients request sources one after the other and can keep them in memory in order to improve the decoding efficiency.

The key novel aspect of SMRA is to study together the storage and transmission issues, while in the current state-of-the-art of information theory, storage and transmission are studied separately. One of the main difficulties of SMRA is to take into account the uncertainty of the successive client's requests, *i.e.* the subset of sources that might be extracted from the stored coded bitstream is random. With this uncertainty, standard coding solutions either have to do re-encoding which leads to a high complexity at the server, or to consider all the possible sets of requests and store all the corresponding encoded bitstreams on a server with a high storage cost. In other words, the problem that is posed by SMRA is the following. Given a set of sources, can we encode them once and only once at a coding rate that minimizes the storage cost, while expecting a rate-efficient extraction of a subset of these sources? Or, if a rate-efficient compression is targeted, how much data should be stored in the bitstream that is *a priori* coded?

A. Overview of the main contributions

Three main contributions are presented in this paper:

- 1) *SMRA definition and new coding scheme*: we formally define the SMRA problem in Section II. We first introduce the new context of SMRA systems, that is a set of sources encoded once for all and stored on a server, and clients asking for subset of sources one after the other. In this case, when a client sends a request to the server at a given time instant, it is assumed that its previous request is still in its memory. A coding scheme for SMRA is then proposed, involving a two-step encoding strategy: *offline encoding* for the storage of sources, and *online extraction* that transmits to the client only the information he needs with respect to his request and without

re-encoding. The problem of minimizing the storage rate (from offline encoding) and the transmission rate (from online extraction) of SMRA is then formulated in the case when the one previously requested source is stored in the client's memory.

- 2) *Theoretical resolution of SMRA problem:* in Section III, we first revisit state-of-the-art results that are related to the SMRA problem. They permit to give insights on the transmission and storage rates that can be achieved by our coding scheme. In Section IV, we then provide a complete information-theoretical study of the transmission and storage rates that can be achieved by SMRA. We, in particular, derive new information-theoretic bounds on these rates for non i.i.d. sources (Section V), no knowledge of source statistics (Sections V-A and V), lossy compression (Section VI), and unbounded memory (Section VII) (not one, but all the previously requested source are stored in the client's memory). In all the considered cases, we show that the minimum achievable storage rate is obtained by taking into account all the possible paths from previous request to current request and is given by the worst one, *i.e.*, the one presenting the lowest correlation between the successively requested sources. In addition, we show that it is possible to construct the stored bitstream so as to achieve a transmission rate equal to the rate needed when re-encoding is possible. We also show that the lack of knowledge on source statistics does not increase the storage rate nor the transmission rate.
- 3) *Practical resolution of SMRA problem:* finally, we propose practical coding schemes that permit to reach these theoretical rates when particular source models are considered. In Section VIII, we explain how incremental channel coding can be used to build computationally efficient algorithms. In Section IX, we consider two particular source models and describe practical implementations of SMRA systems based on Low-Density Parity Check (LDPC)-staircase [3] and Low-Density Parity Check Accumulator (LDPCA) [4] codes.

B. Related problems

The SMRA framework defined in this paper shows connections with other problems that have been studied in the field of multi-source compression. First, in the problem of data delivery with caching, a server makes available a database, and each of the clients involved in the network also stores a limited part of the content of the database. Then, during the delivery phase, the clients request some data to the server, that has to send back the requested data through a single link shared by all the clients. When studying the caching problem, the objective is to minimize the delivery rate that is the amount of data transmitted from the server to the client, under constraints on the client's memory that is the amount of data stored by the clients. For this problem, theoretical rate-memory regions were derived in [5] for i.i.d. sources, and in [6] for correlated sources, where the correlation is between sources. In the SMRA framework, the transmission rate has the same definition as for caching, but the storage rate corresponds to the amount of data stored by the server, while in caching, it corresponds to data stored by the clients. As another difference in the system definition, the caching framework considers only one single common transmission link from the server to all clients, while SMRA assumes direct links from the server to each of the clients. The theoretical approaches also differ, since in the caching formulation, the data is assumed to be already in a compressed form and each item of the database is compressed individually and separately from the others. The main issue of caching is hence to deliver the subset of data that corresponds to the clients' requests with only one common link. Hence the source coding aspects are not discussed in the caching framework while they are the central point of our analysis. In the same way, distributed storage presented in [7] is interested in duplication and data repair in case of server failure, but does not consider the source coding aspects of data storage.

From a source coding perspective, several works show connections with the SMRA framework. First, in source coding with multiple descriptions [8], [9], the source is compressed into different packets such that one packet is sufficient to reconstruct the source, but receiving more packets improves the quality of the reconstruction. The multiple description problem is suboptimal in the sense that the rate needed to reach a given distortion D when several packets are received is higher than the standard rate-distortion function of the source. The suboptimality comes from the constraint that each individual packet must be sufficient to reconstruct the source with a sufficient quality. On the opposite, SMRA corresponds to a problem with one single description from which we want to extract some data that corresponds to the client's request.

The information-theoretic problems that are the closest to SMRA are source coding with multiple decoders [10], and source coding with unknown statistics between the source and the side information [11]. In the SMRA scheme, the side information is the previous request which is assumed to be in the user's memory. Standard problems of source coding with side information either study the storage rate [10] or the transmission rate [11], but none of them considers the joint design of these two quantities. The analysis of these results however gives insights on the transmission and storage rates that can be jointly achieved by SMRA, and this is why we restate them in Section II.

C. Notations

In this paper, a random source X is denoted using uppercase; the source X generates a sequence of random variables denoted X_i using uppercase and index i while its realization x_i is denoted using lowercase; a random vector \mathbf{X} is denoted using boldface uppercase and its realization \mathbf{x} is denoted using boldface lowercase. An n -length vector $\mathbf{X}^n = (X_1, \dots, X_n)$ containing elements X_1 to X_n is denoted using superscript n . The alphabet \mathcal{X} of a random variable is denoted with calligraphic letter, and with the same letter as the random variable. $|\mathcal{X}|$ denotes the cardinality of the set \mathcal{X} . If \mathcal{B} is a finite source alphabet, then \mathcal{B}^+ is the set of all finite strings from \mathcal{B} . $\forall \mathbf{b} \in \mathcal{B}^+$, $|\mathbf{b}|$ stands for the length of \mathbf{b} .

In the case of multiple sources, the set \mathcal{J} of source indexes is denoted using calligraphic letter and the set of sources with indexes in \mathcal{J} is denoted $X_{\mathcal{J}}$. The k^{th} source is then identified with an index inside brackets *i.e.* $_{(k)}$. $\llbracket 1, N \rrbracket$ stands for the set of integers between 1 and N .

II. SEQUENTIAL MASSIVE RANDOM ACCESS PROBLEM

In this section, we define the general *Sequential Massive Random Access* (SMRA) problem. We then describe Free Viewpoint Television (FTV) as a particular instance of SMRA.

A. Graph-based definition of SMRA framework

The SMRA framework is illustrated in Figure 1. Let $\{X_{(\ell)}\}_{1 \leq \ell \leq L}$ be a set of L sources $X_{(\ell)} = (\mathbf{X}_{(\ell)}^n)_{n \rightarrow \infty}$. In the following, the sources can be *general* or *i.i.d.* as defined below.

Definition 1 (Discrete general sources): Let \mathcal{X} be a discrete alphabet. The set $\{X_{(\ell)}\}_{1 \leq \ell \leq L}$ is said to be a *set of L discrete general sources* if the L sources generate random vectors of size $n \geq 1$ according to the joint probability distribution $P(\mathbf{x}_{(1)}^n, \dots, \mathbf{x}_{(L)}^n)$, where $\mathbf{x}_{(1)}^n, \dots, \mathbf{x}_{(L)}^n \in \mathcal{X}^n, \forall n \geq 1$.

Definition 2 (Discrete i.i.d. sources): Let \mathcal{X} be a discrete alphabet. The set $\{X_{(\ell)}\}_{1 \leq \ell \leq L}$ of general sources is said to be a *set of L discrete i.i.d. sources* if the joint probability distribution can be factorized as:

$$\forall n \geq 1, \quad P(\mathbf{x}_{(1)}^n, \dots, \mathbf{x}_{(L)}^n) = \prod_{i=1}^n P(x_{(1),i}, \dots, x_{(L),i}) \quad (1)$$

where $P(x_{(1),i}, \dots, x_{(L),i})$ does not depend on i , and $x_{(\ell),i} \in \mathcal{X}, \forall i \in [1, N], \ell \in [1, L]$.

As shown in Fig. 1, the n symbols of each source are acquired, encoded, and stored as a bitstream on a *server*. These three operations are done *a priori* (i.e., offline), which means that they are done once, without knowing which of the sources will be requested by clients. At the other side of the network, the clients want to sequentially access some of the sources, and thus send consecutive requests to the server.

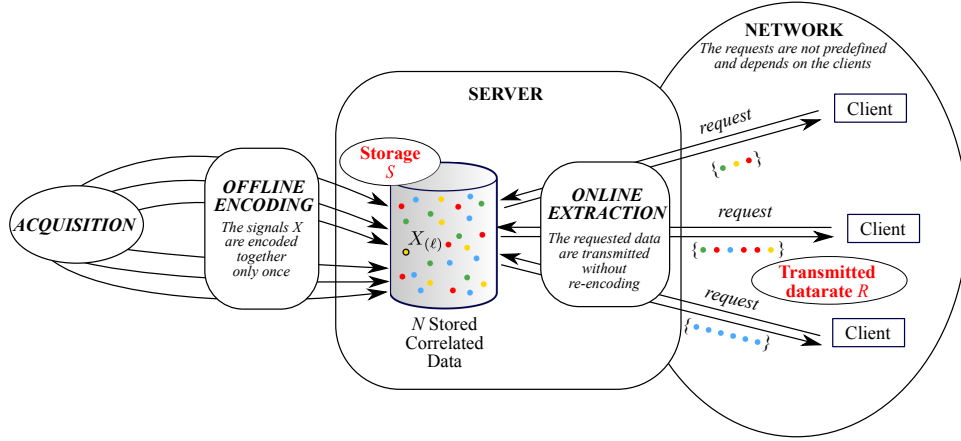


Fig. 1: SMRA framework

Consider a client sequentially asking for $K \leq L$ sources. The request takes the form of a vector \mathbf{K} , such that $\mathbf{K} = [k_0, k_1, \dots, k_t, \dots, k_K]$, where $k_t \in \{1, \dots, L\}$ for $t \geq 1$ and $k_0 = 0$. $X_{(0)}$ is, by convention for initialization, a source with zero entropy. The vector \mathbf{K} indicates that the source of index k_1 is requested first, and that the source of index k_t is requested just after the source of index k_{t-1} for all $1 \leq t \leq K$. Since the successive indices k_t design a path through the source set, we call the request vector \mathbf{K} the *client navigation*. The index t in k_t is called the *navigation instant*. A request to the source of index k_t corresponds to a request to all the n symbols of $\mathbf{x}_{(k_t)}^n$.

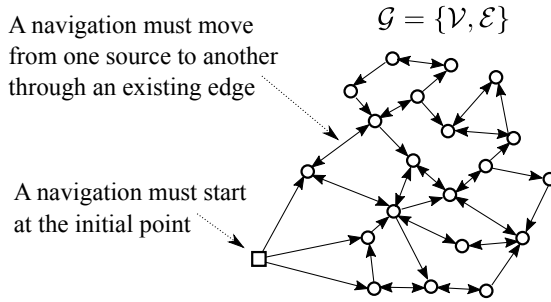


Fig. 2: The possible navigations in the context of SMRA can be represented with a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$. Each vertex is labeled by an integer belonging to $0 \leq \ell \leq L$ and represents a source. An edge in \mathcal{E} from ℓ to ℓ' indicates that a client can have access to source $X_{(\ell')}$, once he has accessed source $X_{(\ell)}$. The graph root represents the source $X_{(0)}$ of zero entropy introduced for initialization and is represented by a square. The L other sources are represented by circles.

In general, the client's navigation is constrained by a structure inherent to the set of sources $\{X_{(\ell)}\}_{1 \leq \ell \leq L}$ and that depends on the application (e.g. FTV). This structure is given *a priori* and can be described by an oriented rooted graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, in which the set of vertices \mathcal{V} represents the $L + 1$ sources (including $X_{(0)}$ corresponding to the root), i.e., $\mathcal{V} = \{0, 1, \dots, L\}$. There is an edge from vertex ℓ to vertex ℓ' if a client is allowed to request the source of index ℓ' just after having requested the source of index ℓ . This is summarized in Figure 2, where the root is depicted as a square.

In the SMRA framework, we assume that the client still has in memory the previously requested source $X_{(k_{t-1})}$ when he receives his

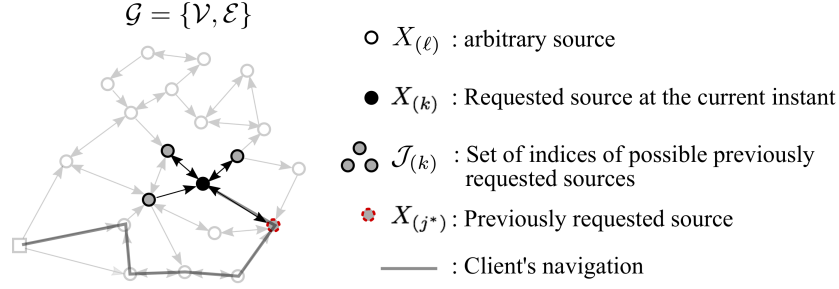


Fig. 3: SMRA notations where the root (vertex 0) is denoted by a square.

current requested source $X_{(k_t)}$ ¹. Therefore, using the graph structure, the encoder knows that if a client requests a source of index k_t with $t > 1$, its previous request k_{t-1} is necessarily a neighbor of vertex k_t in the graph \mathcal{G} . In other words, the processing of source $X_{(k_t)}$ depends only on the sources in its neighborhood. This enables us, in the following, to drop the navigation instant t by considering only two navigation temporalities: the current navigation instant and the previous one. More formally, we assume that, at the *current navigation instant*, a client wants to access the source of index k . $X_{(k)}$ is thus called the *requested source*. At the *previous instant*, the indices of possible *previously requested sources* are the neighbors of vertex k in the graph \mathcal{G} and they are gathered in a set denoted by $\mathcal{J}_{(k)}$. Moreover, the actual previously requested source is denoted by $X_{(j^*)}$ with $j^* \in \mathcal{J}_{(k)}$.

In the next section, we give an interpretation of these notations for one practical example of SMRA system: FTV.

B. Free-viewpoint television: an example of SMRA system

As stated in introduction, FTV [2] is a particular instance of SMRA. In FTV, as illustrated in Figure 4, a client can navigate between different views of a video, which means that at each instant, he picks one frame $X_{(k)}$ (*i.e.*, one image belonging to one view) among all the possible ones $\{X_{(\ell)}\}_{1 \leq \ell \leq L}$. In this model, each frame $X_{(\ell)}$ is seen as a source; each $\mathbf{X}_{(\ell)}^n$ being an image (*i.e.*, a set of pixels). Considering a free viewpoint acquisition system of C cameras acquiring each M successive frames, one obtains a set of $L = C \times M$ sources.

According to the notations of Section II-A, the current requested frame is $X_{(k)}$ (in black in Figure 4) and the client has already requested one frame $X_{(j^*)}$, with $j^* \in \mathcal{J}_{(k)}$ (in grey in Figure 4). The set $\mathcal{J}_{(k)}$ depends on the navigation degree of freedom that is given to the client, which defines the graph \mathcal{G} represented by the arrows in Figure 4. For instance, as an extreme case, the client may be allowed to request any frame from any time of any camera. In this case, the set $\mathcal{J}_{(k)}$ contains all the frame indices $\{0, 1, \dots, L\}$ and the graph \mathcal{G} is fully connected. Alternatively, as in most standard navigation scenarios, the client may only be able to move to one of the next views at each time instant [12]. In the latter case, the set $\mathcal{J}_{(k)}$ consists of three frame indices, as depicted in Figure 4.

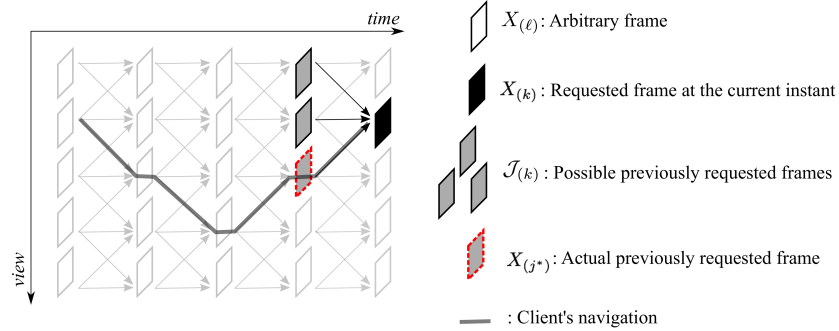


Fig. 4: The free viewpoint navigation is a particular example of SMRA.

C. Coding scheme definition for SMRA

Consider the SMRA problem depicted in Figure 5, where only one source $X_{(k)}$ is to be recovered losslessly. During the first phase, the source $X_{(k)}$ is encoded (compressed) by the *offline encoder* into an index sequence at storage rate $S_{(k)}$ bits per source symbol under the assumptions that (i) all the realizations of the sources in $\mathcal{J}_{(k)}$ are known, (ii) one source from $\mathcal{J}_{(k)}$ will be available at the client's decoder. The index sequence is communicated over a noiseless link to a server. Then, the second phase starts when the client requests the source of index k and specifies the index $j^* \in \mathcal{J}_{(k)}$ of the source already available at its decoder. Upon the client's request, the server extracts an index subsequence at rate $R_{(k)|(j^*)}$ (*online extractor*) and sends it over a noiseless link to the *decoder* which finds an estimate of the requested source. We now define formally the SMRA code and the achievable rates for this code.

Definition 3 (SMRA code): A $((2^{nS_{(k)}}, (2^{nR_{(k)|(j^*)}})_{j^* \in \mathcal{J}_{(k)}})_{1 \leq k \leq L}, n)$ SMRA code for the set of discrete general sources $\{X_{(k)}\}_{1 \leq k \leq L}$ consists, for each source $X_{(k)}$, $1 \leq k \leq L$, of

¹For the sake of clarity, we assume here that the client has only one source in memory: the one requested at the previous instant. The more general case where several requested and memorized sources is studied in Section VII.

- an *offline encoder* $h_{(k)}^{\text{off}}$ that assigns a sequence of $M_{(k)} = |\mathcal{J}_{(k)}|$ indices to each possible set of vectors $(\mathbf{x}_{(k)}^n, (\mathbf{x}_{(j)}^n)_{j \in \mathcal{J}_{(k)}}) \in \mathcal{X}^n \times \mathcal{X}^{n \times M_{(k)}}$

$$h_{(k)}^{\text{off}} : \mathcal{X}^n \times \mathcal{X}^{n \times M_{(k)}} \rightarrow \prod_{m=1}^{M_{(k)}} \{1, \dots, 2^{nr_{k,m}}\} \quad (2a)$$

$$\mathbf{x}_{(k)}^n, (\mathbf{x}_{(j)}^n)_{j \in \mathcal{J}_{(k)}} \mapsto (i_1, \dots, i_{M_{(k)}}) \quad (2b)$$

where $S_{(k)} = r_{k,1} + r_{k,2} + \dots + r_{k,M_{(k)}}$.

- a set of $M_{(k)}$ *online extractors* $h_{(k)|(j^*)}^{\text{on}}$, $j^* \in \mathcal{J}_{(k)}$, that extract a subsequence of indices from the sequence of indices $(i_1, \dots, i_{M_{(k)}})$

$$h_{(k)|(j^*)}^{\text{on}} : \prod_{m=1}^{M_{(k)}} \{1, \dots, 2^{nr_{k,m}}\} \rightarrow \prod_{m \in \mathcal{I}_{(k)|(j^*)}} \{1, \dots, 2^{nr_{k,m}}\} \quad (3a)$$

$$(i_1, \dots, i_{M_{(k)}}) \mapsto (i_m)_{m \in \mathcal{I}_{(k)|(j^*)}} \quad (3b)$$

where $\mathcal{I}_{(k)|(j^*)} \subseteq \{1, \dots, M_{(k)}\}$, and $R_{(k)|(j^*)} = \sum_{m \in \mathcal{I}_{(k)|(j^*)}} r_{k,m} \leq S_{(k)}$

- a set of $M_{(k)}$ *decoders* $g_{(k)|(j^*)}$, $j^* \in \mathcal{J}_{(k)}$, that, given the source realization $\mathbf{x}_{(j^*)}^n$, assigns an estimate $\hat{\mathbf{x}}_{(k)|(j^*)}^n$ to each received subsequence of indices

$$g_{(k)|(j^*)} : \prod_{m \in \mathcal{I}_{(k)|(j^*)}} \{1, \dots, 2^{nr_{k,m}}\} \times \mathcal{X}^n \rightarrow \mathcal{X}^n \quad (4a)$$

$$(i_m)_{m \in \mathcal{I}_{(k)|(j^*)}}, \mathbf{x}_{(j^*)}^n \mapsto \hat{\mathbf{x}}_{(k)|(j^*)}^n \quad (4b)$$

Definition 4 (Achievable storage and transmission rates for a SMRA code): The probability of error for a SMRA code is defined as

$$P_{\text{error}}^n = \mathbb{P} \left(\bigcup_{1 \leq k \leq L} \left(\bigcup_{j^* \in \mathcal{J}_{(k)}} (\hat{\mathbf{x}}_{(k)|(j^*)}^n \neq \mathbf{x}_{(k)}^n) \right) \right) \quad (5)$$

A tuple of rates $((S_{(k)}, (R_{(k)|(j^*)})_{j^* \in \mathcal{J}_{(k)}})_{1 \leq k \leq L})$ is said to be *achievable for SMRA* if there exists a sequence of $((2^{nS_{(k)}}, (2^{nR_{(k)|(j^*)}})_{j^* \in \mathcal{J}_{(k)}})_{1 \leq k \leq L}, n)$ codes such that $\lim_{n \rightarrow \infty} P_{\text{error}}^n = 0$.

The main novelty of the above definition resides in the combination of two encoding mappings: a standard offline encoder that produces the coded source descriptions, and a novel online extractor that can only extract a part of the coded descriptions. The online extractor is a very simple operation introduced because re-encoding is not desirable in massive access to data. The above definition suggests that the encoder and the extractor should be jointly designed in order to minimize both storage $S_{(k)}$ and transmission $R_{(k)|(j^*)}$ rates involved in the definition.

The problem studied in this paper can then be restated as follows: what are the minimum storage $S_{(k)}$ and transmission $R_{(k)|(j^*)}$ rates that can be achieved by SMRA according to the above definition? For sake of clarity, a summary table of the notations introduced in this Section is proposed in Table. I.

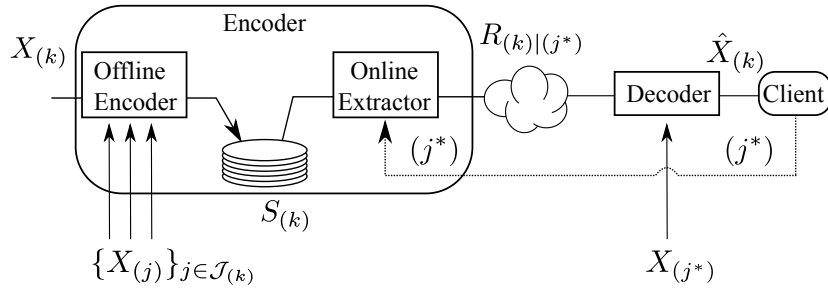


Fig. 5: SMRA: focus on the coding of one source $X_{(k)}$

L	Number of sources
ℓ	Index of an arbitrary source, with $1 \leq \ell \leq L$
k	Index of the source requested by the client, with $1 \leq k \leq L$
$\mathcal{J}_{(k)}$	Set of possible previously requested source indices
j	Index of an arbitrary previously requested source, with $j \in \mathcal{J}_{(k)}$
j^*	Index of the source actually requested by the client at previous instant, with $j^* \in \mathcal{J}_{(k)}$
$S_{(k)}$	Storage rate cost for coding the source of index k
$R_{(k) (j^*)}$	Transmission rate cost for coding the source of index k with source j^* in memory

TABLE I: Notations used for SMRA problem description.

III. REVISITING STATE OF THE ART RESULTS IN THE SMRA FRAMEWORK

SMRA can be seen as a problem of source coding with side information available at the decoder only [13], [14]. This is why existing works in the literature may give partial answers regarding the optimal transmission and storage performance of SMRA. The objective of this section is to review these works and to restate their results according to the definition of the SMRA coding scheme.

A. Source coding with multiple side informations

Sgarro [10] introduces the problem of source coding with multiple decoders, where each decoder has access to a different side information. With our notations (see Table. I), the framework of [10] considers that the source $X_{(k)}$ has to be transmitted to $|\mathcal{J}_{(k)}|$ decoders with different side informations $X_{(j)} \in \mathcal{J}_{(k)}$. Nevertheless, Sgarro considers only one encoding mapping $h_{(k)}^{\text{off}}$ which means that the same index is transmitted to all decoders. Restated in our context, the results of [10] for i.i.d. sources hence give

$$S_{(k)} = \max_{j \in \mathcal{J}_{(k)}} H(X_{(k)} | X_{(j)}) . \quad (6)$$

Further, since no extractor is considered in [10], the above result leads to

$$R_{(k)|(j^*)} = \max_{j \in \mathcal{J}_{(k)}} H(X_{(k)} | X_{(j)}) . \quad (7)$$

These rate expressions may also be derived from other source coding problems. In particular, Draper and Martinian [15] already proposed an analysis of FTV from an information-theoretic perspective. Nevertheless, they do not distinguish between storage and transmission, and they only introduce a single encoding mapping as in [10]. This is why, in our context, the results of [15] (derived for i.i.d. sources) can also be restated as (6). The results of [10] and [15] inspired practical solutions for FTV, see [16] for instance.

In (6) and (7), both the storage and the transmission rates derive from the worst possible side information $X_{(j)}$. These results seem reasonable for the storage rate since the offline encoder does not know which of the sources $X_{(j)} \in \mathcal{J}_{(k)}$ will have been previously requested by the client. However, we will show that the knowledge of the current and previous clients requests can be exploited during data transmission in order to send the data at a rate lower than the worst case.

B. Source coding with side information and unknown statistics

Another problem related to SMRA is source coding with side information at the decoder when the statistics between the source and the side information (*i.e.* their joint probability distribution) are known nor by the encoder nor by the decoder. In [11], [17], in order to deal with the lack of knowledge on the statistics, the encoder successively transmits some pieces of information until the decoder indicates via a feedback link that it was able to successfully decode the message. In our context, the framework of [11], [17] considers that each $X_{(j)} \in \mathcal{J}_{(k)}$ defines one possible joint probability distribution $P(X_{(k)}, X_{(j)})$ between the current source $X_{(k)}$ and the previous request. However, it is worth noting that our setting does not need a feedback link, since, during the on-line phase, the encoders knows the index of the previously requested source (contrary to the off-line phase). The results of [11], [17] indicate that the SMRA coding scheme may be able to achieve

$$R_{(k)|(j^*)} = H(X_{(k)} | X_{(j^*)}) \quad (8)$$

for i.i.d. [11] and stationary [17] sources. Compared to [10], [15], (8) shows that it may be possible to reduce the transmission rate to the conditional entropy of the current request knowing the true previous one. Note that the rate expression of (8) is only a reinterpretation of the results of [11] in our context. In particular, [11] considers only one encoder, and derives only the transmission rate which is shown to vary with the true source statistics.

The rate expressions (6) and (8) give insights on the storage and transmission rates that can be achieved in the context of SMRA. In the following, we first consider i.i.d. sources for simplicity and we show that the storage rate (6) and the transmission rates (8) can be achieved together and are optimal for the SMRA coding scheme of Definition 3. Although we exactly retrieve the expressions (6), (8), deriving the proofs from the beginning permits to introduce a formalism that will be useful to extend the analysis to other cases of interest in practical situations. In the remaining of the paper, we provide extensions to non i.i.d., non stationary sources, to unknown source statistics, and to the case where not only one, but all the previous requests are stored in the client's memory.

IV. OPTIMAL TRANSMISSION AND STORAGE RATES FOR SMRA

A. Discrete i.i.d. sources, known statistics

We first derive the storage and transmission rates for discrete i.i.d. sources, assuming that all the source statistics are known. The proof of Theorem 5 introduces the concepts that allow the generalization to general sources (Section V), without source statistic knowledge (Sections V-A and V), lossy compression (Section VI), and unbounded memory (Section VII).

Theorem 5 (SMRA with previously requested source stored at the decoder): Let $\{X_{(\ell)}\}_{1 \leq \ell \leq L}$ be a set of L discrete i.i.d. sources (see Definition 2). The tuple of rates $((S_{(k)}, (R_{(k)|(j^*)})_{j^* \in \mathcal{J}_{(k)}})_{1 \leq k \leq L})$ is achievable for SMRA (see Definition 4) if and only if $\forall k \in \llbracket 1, L \rrbracket, \forall j^* \in \mathcal{J}_{(k)}$

$$R_{(k)|(j^*)} \geq H(X_{(k)} | X_{(j^*)}) \quad (9a)$$

$$S_{(k)} \geq \max_{j \in \mathcal{J}_{(k)}} H(X_{(k)} | X_{(j)}) \quad (9b)$$

provided that all statistics of the sources are known by both the encoder and the decoder, and that $M_{(k)} = |\mathcal{J}_{(k)}|$ is finite.

Remark: The previously requested source $X_{(j^*)}$ can be seen as a side information partially available at the encoder and perfectly available at the decoder. Indeed, during the offline phase, the encoder only knows the set of possible side informations $X_{(j)}$, $j \in \mathcal{J}_{(k)}$, and their

realizations. Nevertheless, during the online phase, the extractor has access to the exact index (j^*) of the previously requested source available at the decoder as side information.

Remark: When the index 0 of the root node belongs to $\mathcal{J}_{(k)}$, i.e., when the source with index k can be directly accessed, the source needs to be stored at its entropy $H(X_{(k)})$ from (9). However, the source can be transmitted at a lower rate, if $j^* \neq 0$.

Proof: Achievability. In this part, we show the existence of a sequence of SMRA codes that achieves the tuple of rates (9). In what follows, we consider a specific source index k . This source can be requested after one other source has been requested and stored at the decoder. The index of the previously requested source belongs to the set $\mathcal{J}_{(k)}$.

Source ordering. The first step consists in ordering the set of possible previous requests $\mathcal{J}_{(k)}$. More precisely, for each source index k , let

$$\pi_{(k)} : \mathcal{J} \rightarrow \llbracket 1, M_{(k)} \rrbracket \quad (10a)$$

$$j \mapsto \pi_{(k)}(j) \quad (10b)$$

be an ordering mapping of the source indexes in $\mathcal{J}_{(k)}$ such that

$$\begin{aligned} H(X_{(k)} | X_{(\pi_{(k)}^{-1}(1))}) &\leq \dots \leq H(X_{(k)} | X_{(\pi_{(k)}^{-1}(m))}) \\ &\leq \dots \leq H(X_{(k)} | X_{(\pi_{(k)}^{-1}(M_{(k)}))}), \end{aligned} \quad (11)$$

where m stands for the order of a particular source index j : $m = \pi_{(k)}(j)$.

Remark on source ordering. Note that the condition (11) is weak as it only depends on the conditional entropies and can always be obtained. In particular, this condition does not imply that $X_{(k)} \rightarrow X_{(\pi_{(k)}^{-1}(M_{(k)}))} \rightarrow X_{(\pi_{(k)}^{-1}(m))} \rightarrow X_{(\pi_{(k)}^{-1}(1))}$ is a Markov chain and forms a degraded model.

Notation simplification. In what follows, we consider a specific source index k and omit this index to lighten notations.

Coding scheme seen from the offline encoder point of view. As depicted in Figure 6, before sending \mathbf{x}^n , the encoder knows that another sequence (possibly $\mathbf{x}_{(j)}^n$ with $j \in \mathcal{J}$) was sent to the decoder and stored in its memory. This sequence was chosen by the client from the set of possible sequences $\{\mathbf{x}_{(j)}^n\}_{j \in \mathcal{J}}$. The offline encoder knows the set of possible sequences $\{\mathbf{x}_{(j)}^n\}_{j \in \mathcal{J}}$, but does not know which sequence of this set will be available at the decoder, since the encoding is performed offline. Due to similarities with the Slepian Wolf problem [13], we refer to the set of sequences $\{\mathbf{x}_{(j)}^n\}_{j \in \mathcal{J}}$ as the set of possible side informations in the following.

The idea of the proof is to first partition the space \mathcal{X}^n into 2^{nr_1} bins, and then to further partition each bin into 2^{nr_2} bins. The process is repeated M times and this creates an embedded random binning.

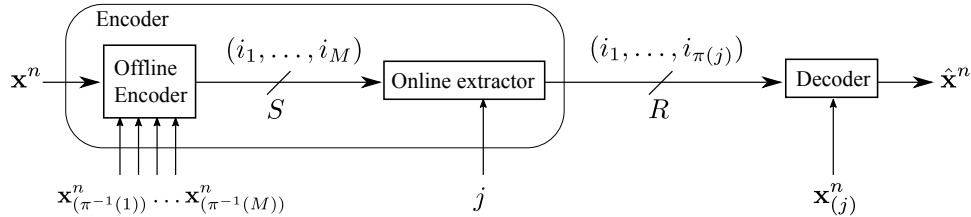


Fig. 6: Coding scheme seen from the offline encoder point of view. $j \in \mathcal{J}$ is a possible previous request available at the decoder. $\pi(j)$ is the order of j according to (11).

Random code generation is performed recursively. First, assign every $\mathbf{x}^n \in \mathcal{X}^n$ to one of 2^{nr_1} bins independently according to a uniform distribution on $\llbracket 1, 2^{nr_1} \rrbracket$. The binning mapping is defined as $f_1(\mathbf{x}^n) = i_1$, where i_1 is the index of the bin assigned to \mathbf{x}^n . Let the bin $\mathcal{B}_{i_1} = f_1^{-1}(i_1)$ denote the preimage of i_1 under the map f_1 .

At the second stage, for each bin index $i_1 \in \llbracket 1, 2^{nr_1} \rrbracket$, assign every $\mathbf{x}^n \in \mathcal{B}_{i_1}$ independently according to a uniform distribution on $\llbracket 1, 2^{nr_2} \rrbracket$. The binning mapping is defined as $f_2(\mathbf{x}^n) = (i_1, i_2)$, where i_1 and i_2 are the indexes assigned to \mathbf{x}^n at the first and second stage respectively. Let the bin $\mathcal{B}_{i_1, i_2} = f_2^{-1}(i_1, i_2)$ denote the preimage of (i_1, i_2) under the map f_2 .

Repeat the sub-partitioning up to stage M . At the last stage, the mapping is $f_M(\mathbf{x}^n) = (i_1, i_2, \dots, i_M)$. Finally, the binning is defined by a set of mappings $\{f_m\}_{1 \leq m \leq M}$, and revealed to both the encoder and the decoder.

Encoding is performed in two parts, see Definition 3 and Figure 6. The *offline encoder* computes a compressed version of the source sequence \mathbf{x}^n as a sequence of bin indexes:

$$h^{\text{off}}(\mathbf{x}^n, (\mathbf{x}_{(j)}^n)_{j \in \mathcal{J}}) = f_M(\mathbf{x}^n) = (i_1, i_2, \dots, i_M),$$

where there are as many bin indexes as the number of possible previous requests. (i_1, i_2, \dots, i_M) corresponds to the index sequence of the bin to which \mathbf{x}^n belongs at the M^{th} and last sub-partitioning. This index sequence is stored in the server.

When the *online extractor* receives the request and, if the index of the source available at the decoder is j , it extracts the $\pi(j)$ first bin indexes from the stored sequence of bin indexes:

$$h_{(j)}^{\text{on}}(i_1, i_2, \dots, i_M) = (i_1, i_2, \dots, i_{\pi(j)})$$

and sends it to the client.

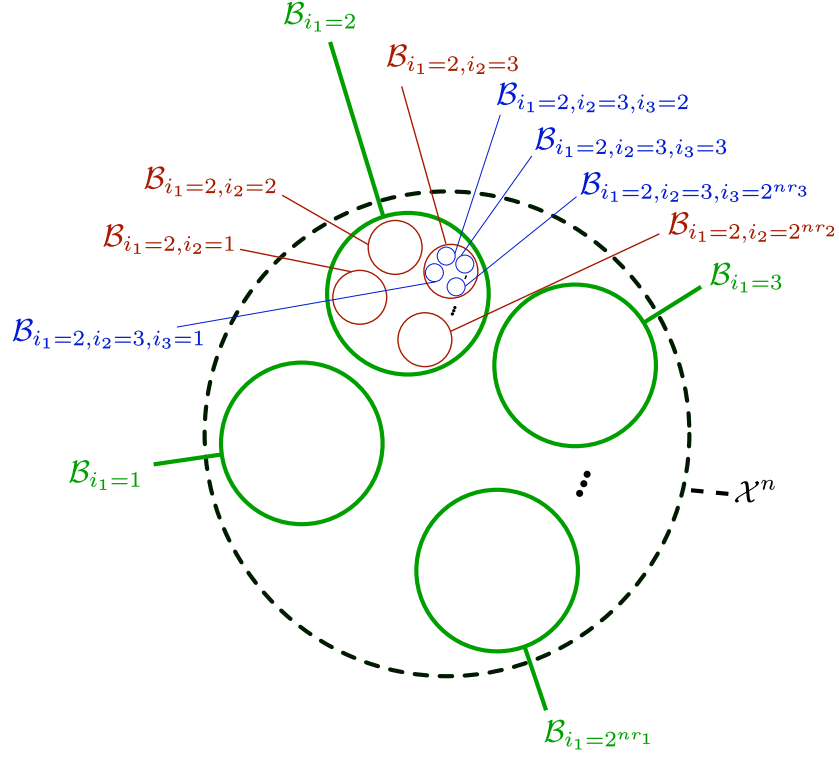


Fig. 7: Random code generation of one source: embedded random binning

Typicality The set of jointly typical sequences, denoted $A_\epsilon^{(n)}(X, X_{(j)})$, is defined as

$$A_\epsilon^{(n)}(X, X_{(j)}) = \left\{ (\mathbf{x}^n, \mathbf{x}_{(j)}^n) \in \mathcal{X}^n \times \mathcal{X}_{(j)}^n : \left| -\frac{1}{n} \log P(\mathbf{x}^n | \mathbf{x}_{(j)}^n) - H(X|X_{(j)}) \right| < \epsilon \right\} \quad (12)$$

Note that this definition of joint typicality differs from [18, Section 7.6]. We further denote $A_\epsilon^n(X|\mathbf{x}_{(j)}^n)$, the set of \mathbf{x}^n sequences jointly typical with $\mathbf{x}_{(j)}^n$. The cardinality of this set can be upper bounded by

$$|A_\epsilon^n(X|\mathbf{x}_{(j)}^n)| \leq 2^{n(H(X|X_{(j)})+\epsilon)} \quad (13)$$

since:

$$1 \geq \sum_{\mathbf{x}'^n \in A_\epsilon^n(X|\mathbf{x}_{(j)}^n)} P(\mathbf{x}'^n | \mathbf{x}_{(j)}^n) \geq |A_\epsilon^n(X|\mathbf{x}_{(j)}^n)| \cdot 2^{-n(H(X|X_{(j)})+\epsilon)} \quad (14)$$

where the last equality follows from (12).

Decoding Given the received index sequence $(i_1, \dots, i_{\pi(j)})$ and the side information $\mathbf{x}_{(j)}$, declare $g((i_1, \dots, i_{\pi(j)}), \mathbf{x}_{(j)}) = \hat{\mathbf{x}}^n = \mathbf{x}^n$ if there is a unique pair of sequences $(\mathbf{x}^n, \mathbf{x}_{(j)}^n)$ such that $f_{\pi(j)}(\mathbf{x}^n) = (i_1, \dots, i_{\pi(j)})$ and $(\mathbf{x}^n, \mathbf{x}_{(j)}^n)$ is jointly typical, denoted $(\mathbf{x}^n, \mathbf{x}_{(j)}^n) \in A_\epsilon^n(X, X_{(j)})$. Otherwise, declare an error.

Probability of error. We now compute the error probability for each possible previously requested source $X_{(j)}$, where $j \in \mathcal{J}$. Let $(\mathbf{X}^n, \mathbf{X}_{(j)}^n) \sim P(\mathbf{x}^n, \mathbf{x}_{(j)}^n)$. We define the events

$$E_0 = \{(\mathbf{X}^n, \mathbf{X}_{(j)}^n) \notin A_\epsilon^n(X, X_{(j)})\} \quad (15)$$

$$E_j = \{\exists \mathbf{x}'^n \neq \mathbf{X}^n : f_{\pi(j)}(\mathbf{x}'^n) = f_{\pi(j)}(\mathbf{X}^n) \text{ and } (\mathbf{x}'^n, \mathbf{X}_{(j)}^n) \in A_\epsilon^n(X, X_{(j)})\}, \quad \forall j \in \mathcal{J} \quad (16)$$

Indeed, there is an error if $(\mathbf{X}^n, \mathbf{X}_{(j)}^n)$ is not typical or if there is another typical sequence pair in the same bin. Then, by the union bound, the error probability is

$$P_{error}^n = \mathbb{P}(E_0 \cup \bigcup_{j \in \mathcal{J}} E_j) \leq \mathbb{P}(E_0) + \sum_{j \in \mathcal{J}} \mathbb{P}(E_j) \quad (17)$$

First, consider E_0 . By the Law of Large Number applied to the i.i.d. random source $-\log P(X|X_{(j)})$, we have that

$$-\frac{1}{n} \sum_{i=1}^n \log P(X_i | X_{(j), i}) \rightarrow H(X|X_{(j)}) \text{ in probability.} \quad (18)$$

Therefore, using similar arguments as for the Asymptotic equipartition property (AEP) [18, Th. 3.1.2], we have $\mathbb{P}(E_0) \rightarrow 0$ as $n \rightarrow \infty$ and hence, $\forall \epsilon > 0$, for n sufficiently large, $\mathbb{P}(E_0) < \epsilon$.

We now compute E_j , $\forall j$

$$\mathbb{P}(E_j) = \sum_{(\mathbf{x}^n, \mathbf{x}_{(j)}^n) \in \mathcal{X}^n \times \mathcal{X}_{(j)}^n} P(\mathbf{x}^n, \mathbf{x}_{(j)}^n) \dots \quad (19)$$

$$\begin{aligned} & \mathbb{P}\left(\exists \mathbf{x}'^n \neq \mathbf{x}^n : f_{\pi(j)}(\mathbf{x}'^n) = f_{\pi(j)}(\mathbf{x}^n) \text{ and } (\mathbf{x}'^n, \mathbf{x}_{(j)}^n) \in A_\epsilon^n\right) \\ & \leq \sum_{(\mathbf{x}^n, \mathbf{x}_{(j)}^n)} P(\mathbf{x}^n, \mathbf{x}_{(j)}^n) \sum_{\substack{\mathbf{x}'^n \neq \mathbf{x}^n \\ (\mathbf{x}'^n, \mathbf{x}_{(j)}^n) \in A_\epsilon^n}} \mathbb{P}\left(f_{\pi(j)}(\mathbf{x}'^n) = f_{\pi(j)}(\mathbf{x}^n)\right) \end{aligned} \quad (20)$$

$$\leq \sum_{(\mathbf{x}^n, \mathbf{x}_{(j)}^n)} P(\mathbf{x}^n, \mathbf{x}_{(j)}^n) \cdot |A_\epsilon^n(X|\mathbf{x}_{(j)}^n)| \cdot 2^{-n(r_1 + \dots + r_{\pi(j)})} \quad (21)$$

$$\leq 2^{-n(r_1 + \dots + r_{\pi(j)})} \cdot 2^{n(H(X|X_{(j)}) + \epsilon)} \quad (22)$$

The probability in (20) is computed over all realizations of the random mapping $f_{\pi(j)}$. The inequality in (20) follows from the fact that there might be many \mathbf{x}'^n satisfying the condition (of the second sum) and from the fact that these events are not disjoint (union bound). For (21), the indexes are chosen according to a uniform distribution, therefore the probability equals $2^{-n(r_1 + \dots + r_{\pi(j)})}$ and $A_\epsilon^n(X|\mathbf{x}_{(j)}^n)$ is the set of \mathbf{x}^n sequences jointly typical with $\mathbf{x}_{(j)}^n$. Finally, since the upper bound of $|A_\epsilon^n(X|\mathbf{x}_{(j)}^n)|$ (13) and the probability $2^{-n(r_1 + \dots + r_{\pi(j)})}$ are independent of $\mathbf{x}^n, \mathbf{x}_{(j)}^n$ and \mathbf{x}'^n , we get (22).

Then, (22) goes to 0 if $r_1 + \dots + r_{\pi(j)} > H(X|X_{(j)}) + \epsilon$. Hence, $\forall j \in \mathcal{J}$ and for sufficiently large n , $\mathbb{P}(E_j) < \epsilon$. Thus the average probability of error P_{error}^n is less than $(M+1)\epsilon$. So, there exists at least one code $\{f_{\pi(j)}\}_j$ with probability of error less than $(M+1)\epsilon$. Thus, if M is finite, we can construct a sequence of codes (indexed by n) with vanishing P_{error}^n .

In other words, for any previously requested source $j \in \mathcal{J}$, there exists a code with vanishing error rate and compression rate $r_1 + \dots + r_{\pi(j)}$, provided that $r_1 + \dots + r_{\pi(j)} \geq H(X|X_{(j)})$ and that M is finite. Finally, to be able to serve any request, the compression scheme needs to prepare a code for any $j \in \mathcal{J}$. Therefore, the storage requires $S \geq \max_{j \in \mathcal{J}} H(X|X_{(j)})$. Moreover, if the actual previous request is j^* , the transmission rate R needs to satisfy $R = r_1 + \dots + r_{\pi(j^*)} \geq H(X|X_{(j^*)})$, which completes the achievability proof.

Converse: outer bound.

- $R_{(k)|(j^*)} \geq H(X_{(k)}|X_{(j^*)})$. Proof by contradiction. If the encoder knows perfectly the realization of the side information $\mathbf{x}_{(j^*)}^n$, and if $R_{(k)|(j^*)} < H(X_{(k)}|X_{(j^*)})$, then, by a conditional version of the lossless (single) source coding theorem [19, Sec. 5.2], the decoder fails to recover the source with probability 1 as $n \rightarrow \infty$. Therefore, if the encoder (compression part) does not know $\mathbf{x}_{(j^*)}^n$, the decoder will also fail.

- $S_{(k)} \geq \max_{j \in \mathcal{J}} H(X_{(k)}|X_{(j)})$. Proof by contradiction.

If $S_{(k)} < \max_{j \in \mathcal{J}} H(X_{(k)}|X_{(j)})$, then with non zero probability, the side information might be $X_{(\bar{j})}$, where $\bar{j} = \arg \max_{j \in \mathcal{J}} H(X_{(k)}|X_{(j)})$. From the converse of the conditional version of the lossless source coding theorem [19, Sec. 5.2], this source can not be recovered at the decoder. ■

B. Non i.i.d. sources, known statistics

In practical situations, sources such as the frames in FTV, or measurements collected by the nodes of a sensor network, present internal non-stationary correlation that needs to be exploited. Such sources correspond to non i.i.d. sources, see Definition 1. As for the i.i.d. setup, a user wants to access a source $X_{(k)}$ among a set of L sources $\{X_{(\ell)}\}_{1 \leq \ell \leq L}$. However, in this case, since the sources are not i.i.d., the joint probability distribution $P(\mathbf{x}_{(1)}^n, \dots, \mathbf{x}_{(L)}^n)$ cannot be factorized as in (1).

For two non i.i.d. sources $X_{(\ell)}$ and $X_{(j)}$, the *spectral conditional entropy* $\bar{H}(X_{(\ell)}|X_{(j)})$ is defined from a lim sup in probability. The lim sup in probability of a sequence $\{A_n\}_{n=1}^{+\infty}$ of random variables is denoted as $\text{p-lim sup}_{n \rightarrow \infty} A_n$ and is given by

$$\text{p-lim sup}_{n \rightarrow \infty} A_n = \inf \left\{ \alpha \mid \lim_{n \rightarrow +\infty} \mathbb{P}(A_n > \alpha) = 0 \right\}. \quad (23)$$

The spectral conditional entropy $\bar{H}(X_{(\ell)}|X_{(j)})$ is then defined as [20, Chapter 7]

$$\bar{H}(X_{(\ell)}|X_{(j)}) = \text{p-lim sup}_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{P(\mathbf{X}_{(\ell)}^n | \mathbf{X}_{(j)}^n)} \quad (24)$$

In (24), the vectors $\mathbf{X}_{(\ell)}^n, \mathbf{X}_{(j)}^n$ are random vectors distributed according to the joint probability distribution $P(\mathbf{x}_{(\ell)}^n, \mathbf{x}_{(j)}^n)$ that can be obtained from $P(\mathbf{x}_{(1)}^n, \dots, \mathbf{x}_{(L)}^n)$. In the above definition, the lim sup in probability is required because general sources are not necessarily ergodic, and hence the random variable $\frac{1}{n} \log \frac{1}{P(\mathbf{X}_{(\ell)}^n | \mathbf{X}_{(j)}^n)}$ does not necessarily converge to a unique value. It can be shown that if the sources $X_{(\ell)}$ and $X_{(j)}$ are i.i.d., then the spectral conditional entropy reduces to the standard definition of entropy, that is $\bar{H}(X_{(\ell)}|X_{(j)}) = H(X_{(\ell)}|X_{(j)})$.

From the definition of the spectral conditional entropy in (24), the rate-storage region can be obtained for non i.i.d. sources by restating Theorem 5 as follows.

Theorem 6: Let $\{X_{(\ell)}\}_{1 \leq \ell \leq L}$ be a set of L general sources (see Definition 1). The tuple of rates $((S_{(k)}, (R_{(k)|(j^*)})_{j^* \in \mathcal{J}_{(k)}})_{1 \leq k \leq L})$ is

achievable for SMRA (see Definition 4) if and only if $\forall k \in \llbracket 1, L \rrbracket, \forall j^* \in \mathcal{J}_{(k)}$

$$R_{(k)|(j^*)} \geq \overline{H}(X_{(k)}|X_{(j^*)}) \quad (25a)$$

$$S_{(k)} \geq \max_{j \in \mathcal{J}_{(k)}} \overline{H}(X_{(k)}|X_{(j)}), \quad (25b)$$

provided that all statistics of the sources are known by both the encoder and the decoder, and that $M_{(k)}$, the size of the set of possible previous requests $\mathcal{J}_{(k)}$ is finite.

Proof: Achievability. In this part, we show the existence of a sequence of codes that achieves the set of rates (25). As for the i.i.d. case, the first step consists in ordering the set of possible previous requests. For each source index k , let

$$\pi : \mathcal{J}_{(k)} \rightarrow \llbracket 1, M_{(k)} \rrbracket \quad (26a)$$

$$j \mapsto \pi(j) \quad (26b)$$

be an ordering mapping of the source indexes in $\mathcal{J}_{(k)}$ such that

$$\begin{aligned} \overline{H}(X_{(k)}|X_{(\pi^{-1}(1))}) &\leq \dots \leq \overline{H}(X_{(k)}|X_{(\pi^{-1}(m))}) \\ &\leq \dots \leq \overline{H}(X_{(k)}|X_{(\pi^{-1}(M_{(k)}))}). \end{aligned}$$

We keep the same notations and conventions as for the proof of Theorem 5. In particular, we omit the index (k) of the source to be encoded. We also rely on the same *random code generation* and *encoding* as for the proof in the i.i.d. case. Indeed, these two steps do not depend on the i.i.d. or non i.i.d. nature of the sources, except for the choice of the rates r_1, \dots, r_M , which will be addressed later in the proof. However, the *decoding* as well as the *error probability analysis* differ from the i.i.d. case, as we now describe.

Decoding. For general sources, define the set $T_\epsilon^n(X, X_{(j)})$ as

$$T_\epsilon^n(X, X_{(j)}) = \left\{ (\mathbf{x}^n, \mathbf{x}_{(j)}^n) \left| \frac{1}{n} \log \frac{1}{P(\mathbf{x}^n|\mathbf{x}_{(j)}^n)} < \sum_{i=1}^{\pi(j)} r_i - \epsilon \right. \right\}. \quad (27)$$

Given the received index sequence $(i_1, \dots, i_{\pi(j)})$ and the side information $\mathbf{x}_{(j)}^n$, declare $\hat{\mathbf{x}}^n = \mathbf{x}^n$ if there is a unique pair of sequences $(\mathbf{x}^n, \mathbf{x}_{(j)}^n)$ such that $f_{\pi(j)}(\mathbf{x}^n) = (i_1, \dots, i_{\pi(j)})$ and $(\mathbf{x}^n, \mathbf{x}_{(j)}^n) \in T_\epsilon^n(X, X_{(j)})$. Otherwise, declare an error.

Probability of error. We define the events

$$E_0 = \{(\mathbf{X}^n, \mathbf{X}_{(j)}^n) \notin T_\epsilon^n(X, X_{(j)})\} \quad (28)$$

$$\begin{aligned} E_j &= \{\exists \mathbf{x}'^n \neq \mathbf{X}^n : f_{\pi(j)}(\mathbf{x}'^n) = f_{\pi(j)}(\mathbf{X}^n) \\ &\quad \text{and } (\mathbf{x}'^n, \mathbf{X}_{(j)}^n) \in T_\epsilon^n(X, X_{(j)})\}, \quad \forall j \in \mathcal{J} \end{aligned} \quad (29)$$

From the union bound, the error probability is given by

$$P_{\text{error}}^n = \mathbb{P}(E_0 \cup \bigcup_{j \in \mathcal{J}} E_j) \leq \mathbb{P}(E_0) + \sum_{j \in \mathcal{J}} \mathbb{P}(E_j) \quad (30)$$

We first consider the error events $E_j, \forall j \geq 1$. The error probability $\mathbb{P}(E_j)$ can be derived as for the i.i.d. case as

$$\mathbb{P}(E_j) = \sum_{(\mathbf{x}^n, \mathbf{x}_{(j)}^n) \in \mathcal{X}^n \times \mathcal{X}_{(j)}^n} P(\mathbf{x}^n, \mathbf{x}_{(j)}^n) \dots \quad (31)$$

$$\begin{aligned} &\mathbb{P}(\exists \mathbf{x}'^n \neq \mathbf{x}^n : f_{\pi(j)}(\mathbf{x}'^n) = f_{\pi(j)}(\mathbf{x}^n) \text{ and } (\mathbf{x}'^n, \mathbf{x}_{(j)}^n) \in T_\epsilon^n) \\ &\leq \sum_{(\mathbf{x}^n, \mathbf{x}_{(j)}^n)} P(\mathbf{x}^n, \mathbf{x}_{(j)}^n) \sum_{\substack{\mathbf{x}'^n \neq \mathbf{x}^n \\ (\mathbf{x}'^n, \mathbf{x}_{(j)}^n) \in T_\epsilon^n}} \mathbb{P}(f_{\pi(j)}(\mathbf{x}'^n) = f_{\pi(j)}(\mathbf{x}^n)) \end{aligned} \quad (32)$$

$$\leq \sum_{(\mathbf{x}^n, \mathbf{x}_{(j)}^n)} P(\mathbf{x}^n, \mathbf{x}_{(j)}^n) \cdot |T_\epsilon^n(X|\mathbf{x}_{(j)}^n)| \cdot 2^{-n \left(\sum_{i=1}^{\pi(j)} r_i \right)} \quad (33)$$

where the set $T_\epsilon^n(X|\mathbf{x}_{(j)}^n)$ is defined as

$$T_\epsilon^n(X|\mathbf{x}_{(j)}^n) = \left\{ \mathbf{x}^n \left| \frac{1}{n} \log \frac{1}{P(\mathbf{x}^n|\mathbf{x}_{(j)}^n)} < \sum_{i=1}^{\pi(j)} r_i - \epsilon \right. \right\}. \quad (34)$$

Here, $|T_\epsilon^n(X|\mathbf{x}_{(j)}^n)|$ can be bounded using the fact that

$$1 \geq \sum_{\mathbf{x}'^n \in T_\epsilon^n(X|\mathbf{x}_{(j)}^n)} P(\mathbf{x}'^n|\mathbf{x}_{(j)}^n) \geq |T_\epsilon^n(X|\mathbf{x}_{(j)}^n)| \exp \left(-n \sum_{i=1}^{\pi(j)} r_i + n\epsilon \right) \quad (35)$$

where $P(\mathbf{x}^n | \mathbf{x}_{(j)}^n) \geq \exp(-n \sum r_i + n\varepsilon)$ comes from the definition of $T_\varepsilon^{(n)}(X | \mathbf{x}_{(j)}^n)$ in (34). Equation (35) gives

$$|T_\varepsilon^{(n)}(X | \mathbf{x}_{(j)}^n)| \leq \exp\left(n \sum_{i=1}^{\pi(j)} r_i - n\varepsilon\right) \quad (36)$$

and as a consequence the error probability $\mathbb{P}(E_j)$ becomes

$$\mathbb{P}(E_j) \leq \exp(-n\varepsilon). \quad (37)$$

This gives $\lim_{n \rightarrow \infty} \mathbb{P}(E_j) = 0$.

Now, according to the definition of $T_\varepsilon^{(n)}(X, X_{(j)})$ in (34), the error probability $\mathbb{P}(E_0)$ can be expressed as

$$\mathbb{P}(E_0) = \mathbb{P}\left(\frac{1}{n} \log \frac{1}{P(\mathbf{X}^n | \mathbf{X}_{(j)}^n)} \geq \sum_{i=1}^{\pi(j)} r_i + \varepsilon\right) \quad (38)$$

If $\sum_{i=1}^{\pi(j)} r_i \geq \bar{H}(X | X_{(j)})$, then from the definition of $\bar{H}(X | X_{(j)})$ in (24), $\mathbb{P}(E_0)$ goes to 0 when n goes to infinity.

From the analysis of $\mathbb{P}(E_0)$ and $\mathbb{P}(E_j)$, and from the expression of P_{error}^n in (30), there exists at least one code $\{f_{\pi(j)}\}_j$ with probability of error less than $(M+1)\varepsilon$. Thus, if M is finite, we can construct a sequence (indexed by n) of codes with $P_{\text{error}}^n \rightarrow 0$.

In other words, for any previously requested source $j \in \mathcal{J}$, there exists a code with vanishing error rate and compression rate $r_1 + \dots + r_{\pi(j)}$, provided that $r_1 + \dots + r_{\pi(j)} \geq \bar{H}(X | X_{(j)})$ and that M is finite. Finally, to be able to serve any request, the compression scheme needs to prepare a code for any $j \in \mathcal{J}$. Therefore, the storage requires $S \geq \max_{j \in \mathcal{J}} \bar{H}(X | X_{(j)})$. Moreover, if the actual previous request is j^* , the transmission rate R needs to satisfy $R = r_1 + \dots + r_{\pi(j^*)} \geq \bar{H}(X | X_{(j^*)})$, which completes the achievability proof.

Converse. The converse is done by contradiction as for the i.i.d. case. In particular, the fact that if $R_{(k)|(j^*)} < \bar{H}(X_{(k)} | X_{(j^*)})$, the decoder fails to recover the source comes from [20, Theorem 7.2.1].

The same result [20, Theorem 7.2.1] shows that if $S_{(k)} < \max_{j \in \mathcal{J}} \bar{H}(X_{(k)} | X_{(j)})$, then with non zero probability, the source indexed by $\bar{j} = \arg \max_{j \in \mathcal{J}} \bar{H}(X_{(k)} | X_{(j)})$ can not be recovered at the decoder. ■

Remark: In the proof of Theorem 5 for i.i.d. sources, the rate condition $R \geq H(X | X_{(j^*)})$ comes from the error events E_j . On the opposite, in the proof of Theorem 6 for general sources, the rate condition $R \geq \bar{H}(X | X_{(j^*)})$ comes from E_0 . This is due to the fact that the decoding is different from one case to another. In the i.i.d. case, the decoding relies on the condition that $(\mathbf{x}^n, \mathbf{x}_{(j^*)}^n) \in A_\varepsilon^n(X, X_{(j^*)})$, the set of jointly typical sequences. But the notion of joint typicality cannot be defined for general sources, and hence in the non i.i.d. case, the decoded sequence must satisfy $(\mathbf{x}^n, \mathbf{x}_{(j^*)}^n) \in T_\varepsilon^n(X, X_{(j^*)})$.

V. OPTIMAL TRANSMISSION AND STORAGE RATES FOR UNIVERSAL SMRA

When considering SMRA with real data such as pictures or videos, it is usually difficult to have access to reliable source statistics prior to encoding, since these statistics heavily depend on the considered data. This is why this section assumes that the source statistics are unknown. Generally speaking, this means that the coding scheme does not know anything about the joint probability distribution of the sources, nor about their marginals or any of the conditional probability distributions. Since the general case may be difficult to characterize, we restrict our attention to two particular cases. Section V-A considers i.i.d. sources, while Section V-B considers non i.i.d. sources with joint distribution defined by an unknown parameter vector.

In both cases, we consider that the source statistics can be inferred from the source realizations. The coding scheme has the same structure as when the source statistics are known, with an offline encoder, an online extractor, and a decoder. The offline encoder now estimates the joint statistics of any possible source pair $(X_{(k)}, X_{(j)})_{j \in \mathcal{J}_{(k)}}$ in the form of a *learning sequence*. The offline encoder uses this learning sequence together with the source sequences to compress the source $X_{(k)}$ into a *data sequence*. Upon the client's request, the online extractor sends a subsequence of both the learning sequence and the data sequence to the decoder that estimates the requested source.

The following definition of universal SMRA codes considers variable length coding because the coding mappings and the rates now depend on the realizations of the source sequences $(\mathbf{x}_{(k)}^n, (\mathbf{x}_{(j)}^n)_{j \in \mathcal{J}_{(k)}})$. In this Definition, the set \mathcal{B} is a finite source alphabet with cardinality greater or equal to 2, and the set \mathcal{B}^+ is the set of all finite strings from \mathcal{B} . In addition, $\forall b \in \mathcal{B}^+$, $|b|$ denotes the length of b .

Definition 7 (Universal SMRA code, storage and transmission rates): Let $\{X_{(k)}\}_{1 \leq k \leq L}$ be a set of discrete general sources. An n -length universal SMRA code consists, for each source $X_{(k)}$, $1 \leq k \leq L$, of a set of mappings $\left(h_{(k)}^{\text{off}}, \left(h_{(k)|(j^*)}^{\text{on}}, g_{(k)|(j^*)}\right)_{j^* \in \mathcal{J}_{(k)}}\right)$, where

- $h_{(k)}^{\text{off}}$ is the *offline encoder* that assigns two sequences of $M_{(k)} = |\mathcal{J}_{(k)}|$ strings each (the first one for learning and the second for the data) for each possible set of vectors $(\mathbf{x}_{(k)}^n, (\mathbf{x}_{(j)}^n)_{j \in \mathcal{J}_{(k)}}) \in \mathcal{X}^n \times \mathcal{X}^{n \times M_{(k)}}$

$$h_{(k)}^{\text{off}} : \mathcal{X}^n \times \mathcal{X}^{n \times M_{(k)}} \rightarrow (\mathcal{B}^+)^{M_{(k)}} \times (\mathcal{B}^+)^{M_{(k)}} \\ \mathbf{x}_{(k)}^n, (\mathbf{x}_{(j)}^n)_{j \in \mathcal{J}_{(k)}} \mapsto (\mathbf{i}'_{(j)})_{j \in \mathcal{J}_{(k)}}, (\mathbf{i}_m)_{m \in \mathcal{J}_{(k)}}$$

- the $h_{(k)|(j^*)}^{\text{on}}, j^* \in \mathcal{J}_{(k)}$, are the $M_{(k)}$ *online sequence extractors*, that extract a string from the sequence of learning strings $(\mathbf{i}'_{(j)})_{j \in \mathcal{J}_{(k)}}$, and a subsequence of strings from the sequence of data strings $(\mathbf{i}_m)_{m \in \mathcal{J}_{(k)}}$

$$h_{(k)|(j^*)}^{\text{on}} : (\mathcal{B}^+)^{M_{(k)}} \times (\mathcal{B}^+)^{M_{(k)}} \rightarrow \mathcal{B}^+ \\ (\mathbf{i}'_{(j)})_{j \in \mathcal{J}_{(k)}}, (\mathbf{i}_m)_{m \in \mathcal{J}_{(k)}} \mapsto \mathbf{i}'_{(j^*)}, (\mathbf{i}_m)_{m \in \mathcal{I}_{(k)|(j^*)}},$$

where $\mathcal{I}_{(k)|(j^*)} \subseteq \{1, \dots, M_{(k)}\}$.

- the $g_{(k)|(j^*)}$, $j^* \in \mathcal{J}_{(k)}$ are the $M_{(k)}$ decoders $g_{(k)|(j^*)}$, $j^* \in \mathcal{J}_{(k)}$, that, given the source realization $\mathbf{x}_{(j^*)}^n$, assign an estimate $\hat{\mathbf{x}}_{(k)|(j^*)}^n$ to each received learning string and to each subsequence of data strings

$$g_{(k)|(j^*)} : \mathcal{B}^+ \times (\mathcal{B}^+)^{|\mathcal{I}_{(k)|(j^*)}|} \times \mathcal{X}^n \rightarrow \mathcal{X}^n$$

$$\mathbf{i}'_{(j^*)}, (\mathbf{i}_m)_{m \in \mathcal{I}_{(k)|(j^*)}}, \mathbf{x}_{(j^*)}^n \mapsto \hat{\mathbf{x}}_{(k)|(j^*)}^n$$

The *storage rate* of a universal SMRA code is defined as the number of stored bits per source sample, when encoding a source of finite length n . It is the sum of two rates:

$$s_{(k)}^n(\mathbf{x}_{(k)}^n, (\mathbf{x}_{(j)}^n)_j) = \frac{1}{n} \left(\sum_{j \in \mathcal{J}_{(k)}} |\mathbf{i}'_{(j)}| + \sum_{m \in \mathcal{J}_{(k)}} |\mathbf{i}_m| \right) \log_2 |\mathcal{B}|. \quad (39)$$

The *transmission rate* $r_{(k)|(j^*)}^{(n)}$ of a universal SMRA code is defined as the number of bits per source sample sent, when encoding a source of finite length n .

$$r_{(k)|(j^*)}^n(\mathbf{x}_{(k)}^n, \mathbf{x}_{(j^*)}^n) = \frac{1}{n} \left(|\mathbf{i}'_{(j^*)}| + \sum_{m \in \mathcal{I}_{(k)|(j^*)}} |\mathbf{i}_m| \right) \log_2 |\mathcal{B}| \quad (40)$$

A. Discrete i.i.d. sources, unknown statistics

The next Theorem derives the storage and transmission rates for discrete i.i.d. sources when the joint source distribution is neither known at the encoder nor at the decoder. In particular, it shows that this lack of knowledge does not increase the transmission and storage rates, provided a universal code is used. Before deriving this result, a set of i.i.d. sources with unknown statistics is formally defined. This allows to define the notion of achievable rate for universal SMRA code for i.i.d. sources.

Definition 8 (Discrete i.i.d. source with unknown statistics): Let \mathcal{X} be a discrete alphabet. Let \mathcal{P} be a set of joint probability mass functions of dimension L defined over \mathcal{X}^L . The set $\{X_{(\ell)}\}_{1 \leq \ell \leq L}$ of L discrete i.i.d. sources is said to be *with unknown statistics* if the joint probability distribution P in (1) belongs to the set \mathcal{P} , and if the set \mathcal{P} is known by the encoder and the decoder, but the probability mass function P is unknown.

Definition 9 (Achievable storage and transmission rates for i.i.d. sources): Let $\{X_{(\ell)}\}_{1 \leq \ell \leq L}$ be a set of L discrete i.i.d. sources with unknown statistics. Given P the unknown probability mass function, the *probability of error* for a universal SMRA code is defined as

$$P_{error}^n(P) = \mathbb{P} \left(\bigcup_{1 \leq k \leq L} \left(\bigcup_{j^* \in \mathcal{J}_{(k)}} \left(\hat{\mathbf{x}}_{(k)|(j^*)}^n \neq \mathbf{x}_{(k)}^n \right) \right) \right) \quad (41)$$

where the event probability is defined with respect to P .

A tuple of rates $((S_{(k)}(P), (R_{(k)|(j^*)}(P))_{j^* \in \mathcal{J}_{(k)}})_{1 \leq k \leq L, P \in \mathcal{P}}$ is said to be *universally achievable for SMRA* if there exists a *universal SMRA code* such that $\forall k, j, j^*, P$,

$$\mathbf{p} - \lim_{n \rightarrow \infty} s_{(k)}^n(\mathbf{X}_{(k)}^n, (\mathbf{X}_{(j)}^n)_j) = S_{(k)}(P) \quad (42a)$$

$$\mathbf{p} - \lim_{n \rightarrow \infty} r_{(k)|(j^*)}^n(\mathbf{X}_{(k)}^n, (\mathbf{X}_{(j)}^n)_j) = R_{(k)|(j^*)}(P) \quad (42b)$$

$$\lim_{n \rightarrow +\infty} P_{error}^n(P) = 0 \quad (42c)$$

where $\mathbf{p} - \lim_{n \rightarrow \infty}$ stands for the limit in probability, and where the event probabilities in the $\mathbf{p} - \lim_{n \rightarrow \infty}$ are defined with respect to P .

Theorem 10: Let $\{X_{(\ell)}\}_{1 \leq \ell \leq L}$ be a set of L discrete i.i.d. sources with unknown statistics (see Definition 8). The tuple of rates $((S_{(k)}(P), (R_{(k)|(j^*)}(P))_{j^* \in \mathcal{J}_{(k)}})_{1 \leq k \leq L, P \in \mathcal{P}}$ is universally achievable for SMRA (see Definition 9) if and only if $\forall k \in \llbracket 1, L \rrbracket, \forall j \in \mathcal{J}_{(k)}, \forall P \in \mathcal{P}$,

$$R_{(k)|(j^*)}(P) \geq H(X_{(k)}|X_{(j^*)}) \quad (43a)$$

$$S_{(k)}(P) \geq \max_{j \in \mathcal{J}_{(k)}} H(X_{(k)}|X_{(j)}) \quad (43b)$$

where the entropies are computed with respect to P and provided that $M_{(k)}$, the size of the set of possible previous requests $\mathcal{J}_{(k)}$ is finite.

Proof: First, compute the joint $p_{\mathbf{x}_{(k)}^n, \mathbf{x}_{(j)}^n}$ and conditional $p_{\mathbf{x}_{(k)}^n | \mathbf{x}_{(j)}^n}$ type [21] of each sequence pair $(\mathbf{x}_{(k)}^n, \mathbf{x}_{(j)}^n)$, $\forall j \in \mathcal{J}_{(k)}$. The number of conditional types is upper bounded by $(n+1)^{|\mathcal{X}|^2}$ [18, Th. 11.1.1], [21] and therefore requires at most $|\mathcal{X}|^2 \log_2(n+1)$ bits to be completely determined. The conditional type class $T_{\mathbf{x}_{(k)}^n | \mathbf{x}_{(j)}^n}$ is the set of all $\mathbf{x}_{(k)}^n$ sequences such that $(\mathbf{x}_{(k)}^n, \mathbf{x}_{(j)}^n)$ belongs to the joint type class $T_{\mathbf{x}_{(k)}^n, \mathbf{x}_{(j)}^n}$. From [21], the cardinality of this set is upper bounded by

$$|T_{\mathbf{x}_{(k)}^n | \mathbf{x}_{(j)}^n}| \leq 2^{nH(p_{\mathbf{x}_{(k)}^n | \mathbf{x}_{(j)}^n})} \quad (44)$$

where $H(p_{\mathbf{x}_{(k)}^n | \mathbf{x}_{(j)}^n})$ is the empirical conditional entropy:

$$H(p_{\mathbf{x}_{(k)}^n | \mathbf{x}_{(j)}^n}) = \frac{1}{n} \sum_{i=1}^n \log p_{\mathbf{x}_{(k)}^n | \mathbf{x}_{(j)}^n}(x_{(k)}, i, x_{(j)}, i) \quad (45)$$

Second, construct the same code as in Theorem 5 where the true distribution is replaced by the type. The learning sequence of the offline encoder $h_{(k)}^{\text{off}}$ consists of all conditional types $\{p_{\mathbf{x}_{(k)}^n | \mathbf{x}_{(j)}^n}\}_{j \in \mathcal{J}_{(k)}}$, which requires $M|\mathcal{X}|^2 \log_2(n+1)$ bits. The data sequence is the index

sequence computed by the code, which requires $\max_j nH(p_{\mathbf{x}_{(k)}^n | \mathbf{x}_{(j)}^n})$ bits.

When the online extractor receives the request, it sends to the decoder the conditional type $p_{\mathbf{x}_{(k)}^n | \mathbf{x}_{(j^*)}^n}$ and the subsequence of length $nH(p_{\mathbf{x}_{(k)}^n | \mathbf{x}_{(j^*)}^n})$.

The decoder reconstructs the code from the conditional type which leads a vanishing error probability as in Theorem 5.

Therefore, the storage and transmission of an n -length realization of the source $X_{(k)}$ requires:

$$ns_{(k)}^n(\mathbf{X}_{(k)}^n, (\mathbf{X}_{(j)}^n)_j) = \max_{j \in \mathcal{J}_{(k)}} nH(p_{\mathbf{x}_{(k)}^n | \mathbf{x}_{(j)}^n}) + M|\mathcal{X}|^2 \log_2(n+1) \quad (46a)$$

$$nr_{(k)|(j^*)}^n(\mathbf{X}_{(k)}^n, (\mathbf{X}_{(j)}^n)_j) = nH(p_{\mathbf{x}_{(k)}^n | \mathbf{x}_{(j^*)}^n}) + |\mathcal{X}|^2 \log_2(n+1) \quad (46b)$$

By the weak law of large numbers [18, Th. 11.2.1] and by continuity of the log function, the empirical entropies converge to the entropies. Moreover, if $M_{(k)}$ is finite, then the contribution of the number of types is negligible compared to the contribution of the index as $n \rightarrow \infty$. Therefore, the set of achievable rates remains unchanged (9). As for the converse, we use the same converse as in the proof of Theorem 5. Indeed, the encoder in Theorem 10 has even less information than in Theorem 5 (statistics are unknown). ■

B. Non i.i.d. parametric sources, unknown statistics

Unfortunately, when the statistics of a non i.i.d. source are unknown, we cannot follow the same reasoning as for i.i.d. sources. Indeed, the proof of Theorem 10 relies on the fact that the encoder can determine the joint type of the sequence pair $(\mathbf{x}, \mathbf{x}_{(j)})$. The notion of type is not always defined for non i.i.d. sources since all the symbols of the sequence may statistically depend from each other. Hence, here, we consider additional assumptions for the non i.i.d. sources, in order to be able to give expressions of the transmission rates and of the storage rate when the sources statistics are unknown. The assumptions we add are general enough to represent a large class of sources. The model we consider corresponds to the Sources without Prior (WP-sources) defined in [22].

Definition 11 (WP-sources): Let Θ be a set of vectors of real-valued vectors of length Q . For all $\theta \in \Theta$, denote by $P_\theta(\mathbf{x}_{(1)}^n, \dots, \mathbf{x}_{(L)}^n)$ a joint probability mass function of L sequences of length n , $(\mathbf{x}_{(1)}^n, \dots, \mathbf{x}_{(L)}^n) \in \mathcal{X}^{L \times n}$. Assume that the function $\mathbf{t} \mapsto P_{\mathbf{t}}(\mathbf{x}_{(1)}^n, \dots, \mathbf{x}_{(L)}^n)$ is continuous over Θ and strictly positive, $\forall n \in \mathbb{N}$, $\forall (\mathbf{x}_{(1)}^n, \dots, \mathbf{x}_{(L)}^n) \in \mathcal{X}^{L \times n}$. The set $\{X_{(\ell)}\}_{1 \leq \ell \leq L}$ is said to be a *set of L WP-sources* if it is a set of L sources that generate random vectors according to one of the joint probability distributions $P_\theta(\mathbf{x}_{(1)}^n, \dots, \mathbf{x}_{(L)}^n)$. The set Θ is known by the encoder and the decoder, but the current parameter θ is unknown.

According to the above definition, the parameter vectors θ do not depend on the sequence length n . As an example, Markov sources can be modeled as WP sources where θ represents the unknown transition probabilities between symbols. Further, we denote $\overline{H}_\theta(X_{(k)} | X_{(j^*)})$ the conditional spectrum entropy defined in (24) evaluated for the conditional distribution $P_\theta(\mathbf{X}_{(k)}^n | \mathbf{X}_{(j)}^n)$ that derives from $P_\theta(\mathbf{X}_{(1)}^n, \dots, \mathbf{X}_{(L)}^n)$.

Definition 12 (Achievable storage and transmission rates for WP sources): Consider a set of L WP sources. The *probabilities of error* for a universal SMRA code for a WP-source are defined $\forall \theta \in \Theta$ as

$$P_{error}^n(\theta) = \mathbb{P} \left(\bigcup_{1 \leq k \leq L} \left(\bigcup_{j^* \in \mathcal{J}_{(k)}} (\hat{\mathbf{X}}_{(k)|(j^*)}^n \neq \mathbf{X}_{(k)}^n) \right) \right) \quad (47)$$

where the event probability is defined with respect to the joint probability distribution P_θ .

A tuple of rates $((S_{(k)}(\theta), (R_{(k)|(j^*)}(\theta))_{j^* \in \mathcal{J}_{(k)}})_{1 \leq k \leq L, \theta \in \Theta})$ is said to be *universally achievable for SMRA* if there exists a *universal SMRA code* such that $\forall k, j, j^*, \theta$,

$$\mathbf{p} - \limsup_{n \rightarrow \infty} s_{(k)}^n(\mathbf{X}_{(k)}^n, (\mathbf{X}_{(j)}^n)_j) \leq S_{(k)}(\theta) \quad (48)$$

$$\mathbf{p} - \limsup_{n \rightarrow \infty} r_{(k)|(j^*)}^n(\mathbf{X}_{(k)}^n, (\mathbf{X}_{(j)}^n)_j) \leq R_{(k)|(j^*)}(\theta) \quad (49)$$

$$\lim_{n \rightarrow \infty} P_{error}^n(\theta) = 0 \quad (50)$$

where the event probability in the $\mathbf{p} - \limsup$ is defined with respect to the joint probability distribution P_θ .

When the parameter vector θ is unknown by the encoder and by the decoder, the transmission rates and the storage rate are given in the following theorem.

Theorem 13: Let $\{X_{(\ell)}\}_{1 \leq \ell \leq L}$ be a set of L WP-sources (see Definition 11). Assume that the encoder can produce Q sequences of estimated parameters $\Theta_q^{(n)}$ such that $\forall q \in \{1, \dots, Q\}$, $\Theta_q^{(n)}$ converges in probability to θ_q . The tuple of rates $((S_{(k)}(\theta), (R_{(k)|(j^*)}(\theta))_{j^* \in \mathcal{J}_{(k)}})_{1 \leq k \leq L, \theta \in \Theta})$ is achievable for SMRA (see Definition 12) if and only if $\forall k \in \llbracket 1, L \rrbracket, \forall j^* \in \mathcal{J}_{(k)}, \forall \theta \in \Theta$,

$$R_{(k)|(j^*)}(\theta) \geq \overline{H}_\theta(X_{(k)} | X_{(j^*)}) \quad (51a)$$

$$S_{(k)}(\theta) \geq \max_{j \in \mathcal{J}_{(k)}} \overline{H}_\theta(X_{(k)} | X_{(j)}), \quad (51b)$$

provided that all statistics of the sources are known by both the encoder and the decoder, and that $M_{(k)}$, the size of the set of possible previous requests $\mathcal{J}_{(k)}$ is finite.

Proof: Throughout the proof, the notation $\Theta_n \xrightarrow{P} \theta$ denotes the convergence in probability of Θ_n to θ . For simplicity, here, we assume that $\theta_q \in [0, 1]$, $\forall q \in \{1, \dots, Q\}$, but the proof can be easily generalized to any parameter $\theta_q \geq 0$.

Learning sequence For each of the Q estimated parameters $\theta_q^{(n)}$ obtained from the sequences $(\mathbf{x}_1^n, \dots, \mathbf{x}_L^n)$, the encoder constructs a learning sequence $\mathbf{s}_q^{u_n}$ of length u_n that contains $\lfloor \theta_q^{(n)} u_n \rfloor$ values 1 and $u_n - \lfloor \theta_q^{(n)} u_n \rfloor$ values 0. The parameter $\theta_q^{(n)}$ can hence be estimated as $\hat{\theta}_q^{(u_n)} = \frac{\sum_{i=1}^{u_n} s_{q,i}}{u_n} = \frac{\lfloor \theta_q^{(n)} u_n \rfloor}{u_n}$, and we denote $\hat{\boldsymbol{\theta}}^{(u_n)} = [\hat{\theta}_1^{(u_n)}, \dots, \hat{\theta}_Q^{(u_n)}]$. For source sequences of length n , the rate needed to store and transmit the learning sequences is given by $\frac{Q u_n}{n}$, since for each of the Q parameters θ_q , there are 2^{u_n} possible learning sequences.

Lemma 14: Assume that $\lim_{n \rightarrow \infty} u_n = +\infty$. Then $\forall q \in \{1, \dots, Q\}$, $\hat{\Theta}_q^{(u_n)} \xrightarrow{P} \theta_q$.

Proof: From $\hat{\theta}_q^{(u_n)} = \frac{\lfloor \theta_q^{(n)} u_n \rfloor}{u_n}$, we get the inequality $\frac{\theta_q^{(n)} u_n - 1}{u_n} \leq \hat{\theta}_q^{(u_n)} \leq \frac{\theta_q^{(n)} u_n + 1}{u_n}$, which implies

$$\left| \hat{\theta}_q^{(u_n)} - \theta_q^{(n)} \right| \leq \frac{1}{u_n} \quad (52)$$

Let $\epsilon > 0$. Then, by marginalization,

$$\begin{aligned} \mathbb{P}(|\hat{\Theta}_q^{(u_n)} - \theta_q| > \epsilon) &= \mathbb{P}(|\hat{\Theta}_q^{(u_n)} - \theta_q| > \epsilon \mid |\Theta_q^{(n)} - \theta_q| > \epsilon) \mathbb{P}(|\Theta_q^{(n)} - \theta_q| > \epsilon) \\ &\quad + \mathbb{P}(|\hat{\Theta}_q^{(u_n)} - \theta_q| > \epsilon \mid |\Theta_q^{(n)} - \theta_q| \leq \epsilon) \mathbb{P}(|\Theta_q^{(n)} - \theta_q| \leq \epsilon). \end{aligned} \quad (53)$$

The fact that $\Theta_q^{(n)}$ converges in probability to θ_q implies that $\forall \epsilon > 0$, $\lim_{n \rightarrow \infty} \mathbb{P}(|\Theta_q^{(n)} - \theta_q| > \epsilon) = 0$. Further, by marginalization with respect to the deterministic condition (52), we get

$$\begin{aligned} \mathbb{P}(|\hat{\Theta}_q^{(u_n)} - \theta_q| > \epsilon \mid |\Theta_q^{(n)} - \theta_q| \leq \epsilon) &= \mathbb{P}(|\hat{\Theta}_q^{(u_n)} - \theta_q| > \epsilon \mid |\Theta_q^{(n)} - \theta_q| \leq \epsilon, |\hat{\Theta}_q^{(u_n)} - \Theta_q^{(n)}| \leq \frac{1}{u_n}) \\ &= \mathbb{P}(|\hat{\Theta}_q^{(u_n)} - \theta_q| > \epsilon \mid |\hat{\Theta}_q^{(u_n)} - \theta_q| \leq \frac{1}{u_n} + \epsilon) \\ &= 0. \end{aligned}$$

As a result, from (53), $\lim_{n \rightarrow \infty} \mathbb{P}(|\hat{\Theta}_q^{(u_n)} - \theta_q| > \epsilon) = 0$, which shows the convergence in probability of $\hat{\Theta}_q^{(u_n)}$ to θ_q . \blacksquare

Encoder and decoder We rely on the same encoders and decoders as in the proof with known statistics, except that the decoding condition is now given by

$$T_\epsilon^{(n)}(X, X_{(j)}) = \left\{ (\mathbf{x}^n, \mathbf{x}_{(j)}^n) \left| \frac{1}{n} \log \frac{1}{P_{\hat{\theta}^{(u_n)}}(\mathbf{x}^n | \mathbf{x}_{(j)}^n)} < \sum_{i=1}^{\pi(j)} r_i - \epsilon \right. \right\}. \quad (54)$$

where $P_{\hat{\theta}^{(u_n)}}(\mathbf{x}^n | \mathbf{x}_{(j)}^n)$ is obtained by marginalization of $P_{\hat{\theta}^{(u_n)}}(\mathbf{x}_1^n, \dots, \mathbf{x}_L^n)$.

From the proof for non i.i.d. sources with known statistics, we show that as $n \rightarrow \infty$,

$$R_{(k)|(j^*)} \geq \bar{H}_{\hat{\Theta}}(X_{(k)} | X_{(j^*)}) + \lim_{n \rightarrow \infty} \frac{Qu_n}{n} \quad (55)$$

$$S_{(k)} \geq \max_{j \in \mathcal{J}_{(k)}} \bar{H}_{\hat{\Theta}}(X_{(k)} | X_{(j)}) + \lim_{n \rightarrow \infty} \frac{Qu_n}{n}. \quad (56)$$

where

$$\bar{H}_{\hat{\Theta}}(X_{(\ell)} | X_{(j)}) = \mathbf{p} - \limsup_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{P_{\hat{\Theta}^{(u_n)}}(\mathbf{X}_{(\ell)}^n | \mathbf{X}_{(j)}^n)}. \quad (57)$$

Note that in the above expression, \mathbf{p} (in the \mathbf{p} - lim sup) comes from the true joint probability distribution $P_{\theta}(\mathbf{x}_1^n, \dots, \mathbf{x}_L^n)$, while $P_{\hat{\Theta}^{(u_n)}}(\mathbf{x}_{(\ell)}^n | \mathbf{x}_{(j)}^n)$ comes from $P_{\hat{\theta}^{(u_n)}}(\mathbf{x}_1^n, \dots, \mathbf{x}_L^n)$.

In order to get that $\hat{\Theta}_q^{(u_n)} \xrightarrow{P} \theta_q$ from Lemma 14, we choose the sequence u_n such that $\lim_{n \rightarrow \infty} u_n = +\infty$. In addition, for the excess rate $\frac{Qu_n}{n}$ to vanish as n goes to infinity, we set $u_n = o(n)$. Thus, in order to prove that unknown statistics do not induce any rate loss, we want to show that

$$\bar{H}_{\hat{\Theta}}(X_{(\ell)} | X_{(j)}) = \bar{H}_{\theta}(X_{(\ell)} | X_{(j)}), \quad (58)$$

In the following, we remove the index ℓ for simplicity. Let

$$A_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n) = \frac{1}{n} \log \frac{1}{P_{\hat{\Theta}^{(u_n)}}(\mathbf{X}^n | \mathbf{X}_{(j)}^n)} \quad (59)$$

$$B_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n) = \frac{1}{n} \log \frac{1}{P(\mathbf{X}^n | \mathbf{X}_{(j)}^n)} \quad (60)$$

$$V_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n) = A_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n) - B_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n) \quad (61)$$

Lemma 15: $\forall \epsilon > 0$, $\lim_{n \rightarrow \infty} \mathbb{P}(|V_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n)| > \epsilon) = 0$ ($V_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n)$ converges in probability to 0).

Proof: The function $\mathbf{t} \mapsto P_{\mathbf{t}}(\mathbf{x}_1^n, \dots, \mathbf{x}_L^n)$ is continuous, which implies that the function $\mathbf{t} \mapsto P_{\mathbf{t}}(\mathbf{x} | \mathbf{x}_{(j)}^n)$ is also continuous, since $P_{\mathbf{t}}(\mathbf{x}_{(j)}^n) > 0$ (by strict positivity of $f_{\mathbf{t}}(\mathbf{x}_1^n, \dots, \mathbf{x}_L^n)$). The function $y \rightarrow \frac{1}{n} \log \frac{1}{y}$ is also continuous. Let $\epsilon_2 > 0$. Then by continuity $\forall n \in \mathbb{N}$, $\forall \mathbf{x}^n, \mathbf{x}_{(j)}^n \in \mathcal{X}^n$ of $\mathbf{t} \mapsto \frac{1}{n} \log \frac{1}{P_{\mathbf{t}}(\mathbf{x}^n | \mathbf{x}_{(j)}^n)}$, $\exists \epsilon_1 > 0$ such that

$$\|\theta_1 - \theta_2\| < \epsilon_1 \Rightarrow \forall n \in \mathbb{N}, \forall \mathbf{x}^n, \mathbf{x}_{(j)}^n \in \mathcal{X}^n, \left| \frac{1}{n} \log \frac{1}{P_{\theta_1}(\mathbf{x}^n | \mathbf{x}_{(j)}^n)} - \frac{1}{n} \log \frac{1}{P_{\theta_2}(\mathbf{x}^n | \mathbf{x}_{(j)}^n)} \right| \leq \epsilon_2 \quad (62)$$

Then, by marginalization,

$$\begin{aligned} \mathbb{P}(|V_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n)| > \epsilon_2) &= \mathbb{P}(|V_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n)| > \epsilon_2 \mid \|\hat{\Theta}^{(u_n)} - \theta\| \geq \epsilon_1) \mathbb{P}(\|\hat{\Theta}^{(u_n)} - \theta\| \geq \epsilon_1) \\ &\quad + \mathbb{P}(|V_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n)| > \epsilon_2 \mid \|\hat{\Theta}^{(u_n)} - \theta\| < \epsilon_1) \mathbb{P}(\|\hat{\Theta}^{(u_n)} - \theta\| < \epsilon_1) \end{aligned} \quad (63)$$

From condition (62), $\mathbb{P}(|V_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n)| > \epsilon_2 \mid \|\hat{\Theta}^{(u_n)} - \theta\| < \epsilon_1) = 0$. From Lemma 14, $\lim_{n \rightarrow \infty} \mathbb{P}(\|\hat{\Theta}^{(u_n)} - \theta\| \geq \epsilon_1) = 0$. As a result, $\lim_{n \rightarrow \infty} \mathbb{P}(|V_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n)| > \epsilon_2) = 0$, which completes the proof of the lemma. ■

Lemma 16: $\mathbb{p} - \limsup_{n \rightarrow \infty} B_n = \mathbb{p} - \limsup_{n \rightarrow \infty} A_n$.

Proof: Let $\bar{\beta}$ be the inf of values β such that $\lim_{n \rightarrow \infty} \mathbb{P}(B_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n) > \beta) = 0$ and let $\bar{\alpha}$ be the inf of values α such that $\lim_{n \rightarrow \infty} \mathbb{P}(A_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n) > \alpha) = 0$.

Let $\epsilon > 0$. By marginalization,

$$\begin{aligned} \mathbb{P}(A_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n) > \bar{\beta} + 2\epsilon) &= \mathbb{P}(A_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n) > \bar{\beta} + 2\epsilon \mid |V_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n)| > \epsilon) \mathbb{P}(|V_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n)| > \epsilon) \\ &\quad + \mathbb{P}(A_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n) > \bar{\beta} + 2\epsilon \mid |V_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n)| \leq \epsilon) \mathbb{P}(|V_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n)| \leq \epsilon) \end{aligned} \quad (64)$$

From Lemma 15, $\lim_{n \rightarrow \infty} \mathbb{P}(|V_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n)| > \epsilon) = 0$. In addition,

$$\mathbb{P}(A_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n) > \bar{\beta} + 2\epsilon \mid |V_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n)| \leq \epsilon) \leq \mathbb{P}(B_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n) > \bar{\beta}) \quad (65)$$

Since $\lim_{n \rightarrow \infty} \mathbb{P}(B_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n) > \beta) = 0$, we get $\lim_{n \rightarrow \infty} \mathbb{P}(A_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n) > \bar{\beta} + 2\epsilon) = 0$, which shows that $\bar{\alpha} \leq \bar{\beta} + 2\epsilon$.

By showing in the same manner that $\lim_{n \rightarrow \infty} \mathbb{P}(A_n(\mathbf{X}^n, \mathbf{X}_{(j)}^n) > \bar{\beta} - 2\epsilon) > 0$, we get that $\bar{\beta} - 2\epsilon \leq \bar{\alpha}$. As a result, $\bar{\beta} - 2\epsilon \leq \bar{\alpha} \leq \bar{\beta} + 2\epsilon$. Since this inequality is true $\forall \epsilon > 0$, we can conclude that $\bar{\beta} = \bar{\alpha}$. ■

From Lemma 16, we can deduce that $\bar{H}_{\hat{\Theta}}(X_{(\ell)} | X_{(j)}) = \bar{H}_{\theta}(X_{(\ell)} | X_{(j)})$, which concludes the proof. ■

The results of Theorem 13 show that when the parameter vector θ is unknown, the storage and transmission rates are the same as if the parameters were known.

VI. OPTIMAL TRANSMISSION AND STORAGE RATES-DISTORTION REGIONS

In this section, we describe lossy source coding for SMRA. We now allow some distortion $d(X_{(k)}, \hat{X}_{(k)})$ between the requested source $X_{(k)}$ and its reconstruction $\hat{X}_{(k)}$. The aim is to determine the minimum storage rate $S_{(k)}(\delta)$ and transmission rate $R_{(k)|(j^*)}(\delta)$ that can be achieved with respect to a distortion constraint δ . This problem is addressed in the simpler case of i.i.d. Gaussian sources.

Definition 17 (i.i.d. Gaussian sources): The set $\{X_{(\ell)}\}_{1 \leq \ell \leq L}$ is said to be a *set of L i.i.d. Gaussian sources* if the joint distribution can be factorized as:

$$P(\mathbf{x}_{(1)}^n, \dots, \mathbf{x}_{(\ell)}^n, \dots, \mathbf{x}_{(L)}^n) = \prod_{i=1}^n P(x_{(1),i}, \dots, x_{(L),i}) \quad (66)$$

and if each pair of symbols $(X_{(k),i}, X_{(j),i})$ is jointly Gaussian with covariance matrix

$$\Sigma_{(k,j)} = \begin{bmatrix} \sigma_k^2 & \rho_{k,j} \sigma_k \sigma_j \\ \rho_{k,j} \sigma_k \sigma_j & \sigma_j^2 \end{bmatrix} \quad (67)$$

Since the sources are Gaussian, we consider a quadratic distortion measure $d(X_{(k)}, \hat{X}_{(k)}) = (X_{(k)} - \hat{X}_{(k)})^2$.

As another difficulty of the lossy source coding problem, the source $X_{(j)}$ that serves as side information for $X_{(k)}$ was also reconstructed with a certain distortion. As a first attempt to address this problem, we reinterpret the SMRA framework as follows. For source $X_{(k)}$, we consider an extended set $\tilde{\mathcal{J}}_{(k)}$ that includes all the possible distorted versions of the sources that can be available at the decoder when $X_{(k)}$ is requested. This increases the size of $\tilde{\mathcal{J}}_{(k)}$ compared to $\mathcal{J}_{(k)}$ since the distortion in the previous request depends on the path that was followed by the user in the navigation graph. As a result, each of the possible previous requests given by the navigation graph will appear several times in $\tilde{\mathcal{J}}_{(k)}$, with different distortion levels.

Although the set $\tilde{\mathcal{J}}_{(k)}$ may be difficult to characterize in practical situations, we consider it here in order to provide insights on the tradeoff between transmission-storage rate versus distortion in the lossy context. In addition, the Gaussian model of Definition 17 is hereafter assumed to be between $X_{(k)}$ and the distorted sources contained in $\tilde{\mathcal{J}}_{(k)}$. As a result, the parameters σ_j^2 and $\rho_{k,j}$ will depend on the distortion levels in the $X_{(j)} \in \tilde{\mathcal{J}}_{(k)}$. Although not completely rigorous, we consider this assumption here in order to obtain a first characterization of the rate-storage tradeoff in SMRA.

Under the above assumptions, the transmission-storage rates versus distortion regions for Gaussian sources for SMRA are given in Theorem 18.

Theorem 18: Let $\{X_{(\ell)}\}_{1 \leq \ell \leq L}$ be a set of L i.i.d. Gaussian sources (see Definition 17). For a given parameter δ , the tuple of rates $((S_{(k)}(\delta), (R_{(k)|(j^*)}(\delta))_{j^* \in \tilde{\mathcal{J}}_{(k)}})_{1 \leq k \leq L})$ is achievable for SMRA if and only if $\forall k \in \llbracket 1, L \rrbracket, \forall j^* \in \tilde{\mathcal{J}}_{(k)}$

$$\begin{aligned} \frac{R_{(k)|(j^*)}(\delta)}{S_{(k)}(\delta)} &\geq \frac{R_{k,j^*}(D_{k,j^*})}{\max_{j \in \tilde{\mathcal{J}}_{(k)}} R_{k,j}(D_{k,j})} \\ E[(\hat{X}_{(k)} - X_{(k)})^2 | X_{(j^*)}] &\leq D_{k,j^*} \end{aligned} \quad (68)$$

where

$$R_{k,j}(D_{k,j}) = \begin{cases} \frac{1}{2} \log_2 \frac{\sigma_{k,j}^2}{D_{k,j}} & \text{if } \sigma_{k,j}^2 > D_{k,j}, \\ 0 & \text{otherwise,} \end{cases} \quad (69)$$

with $D_{k,j} = \frac{\delta \sigma_{k,j}^2}{\sigma_{k,j}^2 + \delta \rho_{k,j}^2}$ and $\sigma_{k,j}^2 = \sigma_k^2(1 - \rho_{k,j}^2)$, provided that all statistics of the sources are known by both the encoder and the decoder, and that $M_{(k)}$, the size of the set $\tilde{\mathcal{J}}_{(k)}$ is finite.

Remark: The parameter δ is given by $\delta = E[(\hat{X}_{(k)} - X_{(k)})^2]$ when no previous request is available at the decoder. When a previous source $X_{(j)}$ is available (in principle a distorted version of a possible previous request), the source $X_{(k)}$ can be reconstructed with a distortion

$D_{k,j} \leq \delta$ that depends on the parameter δ and on the source statistics $\rho_{k,j}$, σ_k , σ_j . The parameter δ hence leads to different distortions $D_{k,j}$, depending on the available source $X_{(j)}$. This fact was also observed in [23], [24] where side informations with different statistics are available at different decoders.

Remark: In (68), the transmission rate $R_{(k)|(j^*)}(\delta) = \frac{1}{2} \log_2 \frac{\sigma_{k,j^*}^2}{D_{k,j^*}}$ corresponds to the Wyner-Ziv rate-distortion function for a given target distortion D_{k,j^*} when $X_{(j^*)}$ is the only possible side information. The storage rate $S_{(k)}(\delta)$ is still given by the worst possible rate-distortion function $\max_{j \in \tilde{\mathcal{J}}_{(k)}} R_{k,j}(D_{k,j})$. The tradeoff between the transmission-storage rates and the distortions is addressed through the parameter δ that affects all the distortion levels $D_{k,j}$.

Proof: Source ordering. For each source index k , let

$$\pi : \tilde{\mathcal{J}}_{(k)} \rightarrow \llbracket 1, M_{(k)} \rrbracket \quad (70a)$$

$$j \mapsto \pi(j) \quad (70b)$$

be an ordering of the source indexes in $\tilde{\mathcal{J}}_{(k)}$ such that $m = \pi(j)$ and

$$\rho_{k,\pi^{-1}(1)} \leq \dots \leq \rho_{k,\pi^{-1}(m)} \leq \dots \leq \rho_{k,\pi^{-1}(|\tilde{\mathcal{J}}_{(k)}|)}. \quad (71)$$

We keep the same notations and conventions as for the proof of Theorem 5. In particular, we omit the index (k) of the source to be encoded. The proof involves two proof techniques: Gaussian test channels and typical sequences. This enables to derive the rate-distortion functions in the simpler Gaussian case, but also to deal with incremental coding that is used to transmit the codeword from the encoder to the decoder.

Achievability : Random code generation. Let U be a random variable such that $U = \gamma X + \Phi$, where $\Phi \sim \mathcal{N}(0, \gamma\delta)$ is independent of X and $\gamma = 1 - \frac{\delta}{\sigma_k^2}$. This test channel comes from [11]. Generate 2^{nr_0} sequences \mathbf{U}^n at random, where $r_0 = I(X; U) + \epsilon$. Denote by \mathcal{C} the set of generated sequences $\mathbf{u}^n(s)$ and index them with $s \in \llbracket 1, 2^{nr_0} \rrbracket$. Assign each $\mathbf{u}^n \in \mathcal{C}$ to incremental bins, following the same process as in the proof of Theorem 5. The size of the M successive bins depends on the values r_i such that $r_1 + \dots + r_{\pi(j)} = I(X; U) - I(X_{(j)}; U) + 5\epsilon$. It can be verified that $\forall j$, $r_{\pi(j)} \geq 0$ (condition on successive $\rho_{k,j}$ of Theorem 18). This defines mappings $f_{\pi(j)}(\mathbf{u}^n) = (i_1, \dots, i_{\pi(j)})$ where the $i_{\pi(j)}$ are the indices of the successive bins to which \mathbf{u}^n belongs.

Encoding. Given a sequence \mathbf{x}^n , find a sequence $\mathbf{u}^n(s) \in \mathcal{C}$ such that $(\mathbf{x}^n, \mathbf{u}^n(s)) \in A_\epsilon^n(X, U)$. The offline encoder then sends to the storage unit the index sequence i_1, \dots, i_M to which $\mathbf{u}^n(s)$ belongs. Upon request of the source X and previous request j , the online extractor sends to the user the index sequence $(i_1, \dots, i_{\pi(j)})$ to which $\mathbf{u}^n(s)$ belongs.

Decoding. Given the received index sequence $(i_1, \dots, i_{\pi(j)})$ and the side information $\mathbf{x}_{(j)}^n$, declare $\hat{\mathbf{u}}^n = \mathbf{u}^n(s)$ if there is a unique pair of sequences $(\mathbf{u}^n(s), \mathbf{x}_{(j)}^n)$ such that $f_{\pi(j)}(\mathbf{u}^n(s)) = (i_1, \dots, i_{\pi(j)})$ and $(\mathbf{u}^n(s), \mathbf{x}_{(j)}^n) \in A_\epsilon^n(U, X_{(j)})$, where A_ϵ^n is defined in (12). Then compute $\hat{X} = \alpha U + \beta X_{(j)}$, with $\alpha = \frac{\sigma_{k,j}^2}{\gamma\sigma_{k,j}^2 + \delta}$ and $\beta = \frac{\rho_{k,j}\sigma_k}{\sigma_j} \frac{\delta}{\gamma\sigma_{k,j}^2 + \delta}$.

Probability of error. We define the error events:

$$E_{0,0} = \{(\mathbf{X}^n, \mathbf{X}_{(j)}^n) \notin A_\epsilon^n(X, X_{(j)})\} \quad (72)$$

$$E_{0,1} = \{(\mathbf{X}^n, \mathbf{X}_{(j)}^n) \in A_\epsilon^n(X, X_{(j)}) \\ \text{but } \nexists s \text{ such that } (\mathbf{X}^n, \mathbf{u}^n(s)) \in A_\epsilon^n(X, U)\} \quad (73)$$

$$E_{0,2} = \{(\mathbf{X}^n, \mathbf{u}^n(s)) \in A_\epsilon^n(X, U) \\ \text{but } \nexists s \text{ such that } (\mathbf{X}_{(j)}^n, \mathbf{u}^n(s)) \in A_\epsilon^n(X_{(j)}, U)\} \quad (74)$$

$$E_j = \{\exists s' \neq s : f_{\pi(j)}(\mathbf{u}^n(s')) = f_{\pi(j)}(\mathbf{u}^n(s)) \\ \text{and } (\mathbf{X}_{(j)}^n, \mathbf{u}^n(s')) \in A_\epsilon^n(X_{(j)}, U)\}, \quad \forall j \in \mathcal{J} \quad (75)$$

First, using the same arguments as in the proof of Theorem 5, we have $\mathbb{P}(E_{0,0}) \rightarrow 0$ as $n \rightarrow \infty$. Second, $\mathbb{P}(E_{0,1}) \rightarrow 0$ as $n \rightarrow \infty$ by the source coding theorem and since $r_0 > I(U; X)$. Then, $\mathbb{P}(E_{0,2}) \rightarrow 0$ as $n \rightarrow \infty$ by the Markov Lemma. We now evaluate E_j , $\forall j \in \mathcal{J}$

$$\mathbb{P}(E_j) = \sum_{(\mathbf{u}^n, \mathbf{x}_{(j)}^n) \in \mathcal{C} \times \mathcal{X}_{(j)}^n} P(\mathbf{u}^n, \mathbf{x}_{(j)}^n) \dots \quad (76)$$

$$\begin{aligned} & \mathbb{P}(\exists s' \neq s : f_{\pi(j)}(\mathbf{u}^n(s')) = f_{\pi(j)}(\mathbf{u}^n(s)) \text{ and } (\mathbf{u}^n(s'), \mathbf{x}_{(j)}^n) \in A_\epsilon^n) \\ & \leq \sum_{(\mathbf{u}^n, \mathbf{x}_{(j)}^n)} P(\mathbf{u}^n, \mathbf{x}_{(j)}^n) \sum_{\substack{s' \neq s \\ f_{\pi(j)}(\mathbf{u}^n(s')) = f_{\pi(j)}(\mathbf{u}^n(s))}} \mathbb{P}((\mathbf{u}^n(s'), \mathbf{x}_{(j)}^n) \in A_\epsilon^n) \\ & \leq 2^{-n(r_0 - (r_1 + \dots + r_{\pi(j)}))} \cdot 2^{n(I(U; X_{(j)})) - 3\epsilon} \end{aligned} \quad (77)$$

Hence $\forall j \in \mathcal{J}$, $\mathbb{P}(E_j) \rightarrow 0$ as $n \rightarrow \infty$ since we have $r_0 - (r_1 + \dots + r_{\pi(j)}) < I(U; X_j) - 3\epsilon$. At the end, $P_{error}^n = \mathbb{P}(E_{0,0} \cup E_{0,1} \cup E_{0,2} \cup \bigcup_{j \in \mathcal{J}} E_j) \rightarrow 0$ as $n \rightarrow \infty$.

Distortion. With the above parameters α , β , and δ , the actual distortion can be calculated as

$$\begin{aligned} D_j &= E[(\hat{X} - X)^2 | X_{(j)}] \\ &= (1 - P_{error}^n) \frac{\delta\sigma_{k,j}^2}{\sigma_{k,j}^2 + d\rho_{k,j}^2} + P_{error}^n \delta_{max} \end{aligned} \quad (79)$$

where δ_{max} is a constant that represents the maximum possible distortion. As a result, $D_j \rightarrow \frac{\delta \sigma_{k,j}^2}{\sigma_{k,j}^2 + \delta \rho_{k,j}^2}$ as $n \rightarrow \infty$.

In other words, for any previously requested source $j \in \mathcal{J}$, there exists a code with a distortion $D_j = \frac{\delta \sigma_{k,j}^2}{\sigma_{k,j}^2 + \delta \rho_{k,j}^2}$ and compression rate $r_1 + \dots + r_{\pi(j)}$, provided that $r_1 + \dots + r_{\pi(j)} \geq I(X; U) - I(U; X_{(j)}) = \frac{1}{2} \log_2 \frac{\sigma_{k,j}^2}{D_j}$ and that M is finite. Finally, to be able to serve any request j with distortion D_j , the compression scheme needs to prepare a code for any $j \in \mathcal{J}$. Therefore, the storage requires $S \geq \max_{j \in \mathcal{J}} \frac{1}{2} \log_2 \frac{\sigma_{k,j}^2}{D_j}$. Moreover, if the actual previous request is j^* , to achieve the distortion D_{j^*} , the transmission rate R needs to satisfy $R \geq -\frac{1}{2} \log_2 \frac{\sigma_{k,j^*}^2}{D_{j^*}}$, which completes the achievability proof.

Converse: The converse can be done by contradiction as in the lossless cases. ■

Remark: Theorem 18 still holds when the source statistics (the variances and the correlation coefficient) are unknown. The lack of knowledge on the joint statistics between \mathbf{u}^n and $\mathbf{x}_{(j)}^n$ can be handled by the method of types, as described in the proof of Theorem 10. The parameters σ_k and σ_j can be estimated online at the encoder and at the decoder, respectively. The parameter $\rho_{k,j}$ can be estimated from the relation $E[uy] = \gamma \rho_{k,j} \sigma_k \sigma_j$.

VII. GENERALIZATION TO UNBOUNDED MEMORY AT THE DECODER

In Sections IV to VI, one considers that only the previously requested source is stored in the decoder memory. In this section, Theorem 20 provides the gain that can be obtained when all previously requested sources are stored at the decoder.

Definition 19 (Set of possible previous requests): Let k denote the current request. The *set of possible previous requests* $\mathcal{F}_{(k)} = \{\mathcal{E}_1, \dots, \mathcal{E}_{M_{(k)}}\}$ is a set of subsets of $\llbracket 0, L \rrbracket / \{k\}$, where each subset contains the previous requests that might be available at the decoder. $M_{(k)} = |\mathcal{F}_{(k)}|$.

Example: To illustrate $\mathcal{F}_{(k)}$ on a simple example, consider the directed graph in Figure 8. Here, $\mathcal{F}_{(2)} = \{\{0\}, \{0, 1\}\}$, and $\mathcal{F}_{(3)} = \{\{0, 1\}, \{0, 2\}, \{0, 1, 2\}\}$. Note that the graph is directed giving rise to ordered sequences, but the subsets are specified without any order, as the order does not matter in the conditional entropy. Therefore different paths in the directed graph may lead to the same subset in $\mathcal{F}_{(k)}$.

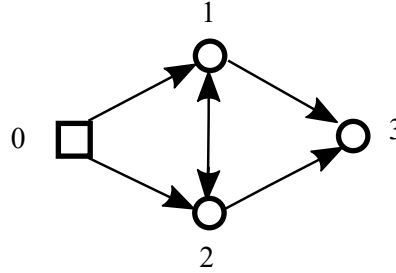


Fig. 8: Graph example, where 0 stands for the initial node. $\mathcal{J}_{(1)} = \{0, 2\}$ and $\mathcal{J}_{(3)} = \{1, 2\}$.

Theorem 20 (SMRA while storing all the previously requested sources at the decoder): Let $\{X_{(\ell)}\}_{1 \leq \ell \leq L}$ be a set of L discrete i.i.d. sources. The tuple of rates $((S_{(k)}, (R_{(k)|(j^*)})_{j^* \in \mathcal{J}_{(k)}})_{1 \leq k \leq L})$ is achievable for SMRA if and only if $\forall k \in \llbracket 1, L \rrbracket, \forall \mathcal{E}^* \in \mathcal{F}_{(k)}$

$$R_{(k)|\mathcal{E}^*} \geq H(X_{(k)}|X_{\mathcal{E}^*}), \quad (80a)$$

$$S_{(k)} \geq \max_{\mathcal{E} \in \mathcal{F}_{(k)}} H(X_{(k)}|X_{\mathcal{E}}). \quad (80b)$$

provided that all statistics of the sources are known by both the encoder and the decoder, that all the previously requested sources are stored at the decoder, and that $|\mathcal{F}_{(k)}|$ is finite.

Remark: Theorem 20 allows one to compare the case where only one single source j^* is known at the decoder (Theorem 5), with the more general case of encoding along an arbitrary path of the navigation graph, where many sources \mathcal{E}^* are available at the decoder (Theorem 20). As intuition suggests, taking into account more sources at the decoder, allows to reduce the rate at which $X_{(k)}$ is stored and transmitted since:

$$H(X_{(k)}|X_{\mathcal{E}^*}) \leq H(X_{(k)}|X_{(j^*)}), \quad (81a)$$

$$\max_{\mathcal{E} \in \mathcal{F}_{(k)}} H(X_{(k)}|X_{\mathcal{E}}) \leq \max_{j \in \mathcal{J}_{(k)}} H(X_{(k)}|X_{(j)}). \quad (81b)$$

where $j^* \in \mathcal{E}^*$ is the request preceding k and $\mathcal{J}_{(k)}$ (defined in Section II-A) is the one-hop neighborhood of the source with index k . Equality in (81a) and (81b) occurs if all the nodes of the one-hop neighborhood $\mathcal{J}_{(k)}$ are all connected to the initial node 0.

Proof: Let Y be a sufficient statistic of $X_{\mathcal{E}^*}$ for $X_{(k)}$, such that

$$H(X_{(k)}|Y) = H(X_{(k)}|X_{\mathcal{E}^*}). \quad (82)$$

Let us apply Theorem 5 with Y as the previous source available at the decoder. Then, applying the conditional entropy equality (82) leads to (80).

Note that, by applying Theorem 7, we also construct, for each source of index k , a reordering of the family of subsets of previous requests (Definition 19) according to

$$H(X_{(k)}|X_{\mathcal{E}_1}) \leq \dots \leq H(X_{(k)}|X_{\mathcal{E}_{|\mathcal{F}_{(k)}|}}). \quad (83)$$

We now show (81). By construction, the previously requested source $X_{(j^*)}$ belongs to the set of previously requested sources $X_{\mathcal{E}^*}$. So, since conditioning reduces entropy, we have (81a).

Moreover, we have

$$\forall \mathcal{E} \in \mathcal{F}_{(k)}, \forall \ell \in \mathcal{E}, H(X_{(k)}|X_{\mathcal{E}}) \leq H(X_{(k)}|X_{(\ell)}), \quad (84)$$

which holds in particular for any source that belongs to \mathcal{E} and to the one-hop neighborhood of $X_{(k)}$, i.e. it holds for any source indexed by $\ell \in \mathcal{E} \cap \mathcal{J}_{(k)}$. Note that the set $\mathcal{E} \cap \mathcal{J}_{(k)}$ may contain many indexes since each set \mathcal{E} corresponds to eventually many paths, as a path is ordered whereas the set is not. Finally, we get

$$\max_{\mathcal{E} \in \mathcal{F}_{(k)}} H(X_{(k)}|X_{\mathcal{E}}) \leq \max_{\mathcal{E} \in \mathcal{F}_{(k)}} \max_{\ell \in \mathcal{E} \cap \mathcal{J}_{(k)}} H(X_{(k)}|X_{(\ell)}) \quad (85)$$

$$= \max_{\ell \in \mathcal{J}_{(k)}} H(X_{(k)}|X_{(\ell)}) \quad (86)$$

where the second equality follows from the fact that $\cup_{\mathcal{E} \in \mathcal{F}_{(k)}} (\mathcal{E} \cap \mathcal{J}_{(k)}) = \mathcal{J}_{(k)}$ i.e. the multi-hop neighborhood of a node k reduced to one-hop is exactly the one-hop neighborhood $\mathcal{J}_{(k)}$. This shows (81b).

Case of equality in (81b). If the nodes of the one-hop neighborhood $\mathcal{J}_{(k)}$ are all connected to the initial node 0 (meaning that the navigation can start at any source indexed by $\mathcal{J}_{(k)}$), then the equality holds in (85) and therefore in (81b). The equality is also achieved if the node in the one-hop neighborhood $\mathcal{J}_{(k)}$ that achieves the maximal entropy in (86) is connected to node 0. ■

VIII. PRACTICAL SCHEME

The encoding and decoding processes detailed in Section IV cannot be used in practice as they rely on random code generation and typical decoding. However, we can construct practical codes that approach the rate-storage bounds in the case of discrete i.i.d. sources, when the correlation between the sources is modeled as an additive channel. This is a consequence of the following facts:

- (i) *Information theoretical duality between common broadcast channel and universal source coding with side information at the receiver:* [15] shows that a code optimal for common broadcasting [25] [26, Chap. 4] (a common information is to be sent to a set of receivers with different channel capacities and a receiver can quit when it receives the information needed to reconstruct the data) can be turned into an optimal code for source coding with side information and feedback link to stop the transmission.
- (ii) *Information theoretical duality between universal source coding with side information at the receiver and SMRA:* Section IV shows that this same code [25] [26, Chap. 4] can be used for SMRA, but with a novel way to determine the transmitted rate, computed at the encoder directly and based on the source ordering defined in the proof of Theorem 5.
- (iii) *Existence of practical codes for the common broadcast channel:* practical channel coding schemes have been proposed for the common broadcast channel and are usually referred to as Fountain codes [27], [28] or rateless codes. These channel codes have the ability to compute repair symbols on the fly in order to meet the channel capacity constraint of each receiver.
- (iv) *Channel codes can be turned into Slepian-Wolf source codes [13]:* one can use the parity bits of a channel block code to perform a lossless compression with side information at the decoder [14] [29, Chap. 6].

We now detail this approach in the context of SMRA with discrete i.i.d. sources. First, the correlations between the source to be encoded $X_{(k)}$ and the side informations $X_{(j)}, j \in \mathcal{J}_{(k)}$ is modeled as a set of channels with the same input $X_{(k)}$ and output $X_{(j)}$ but with different statistics. Then, an incremental and systematic channel code is constructed for this set of channels with rates $\{R_{\pi(j)}^{ch}\}_{j \in \mathcal{J}_{(k)}}$, where $R_{\pi(j)}^{ch}$ stands for the ratio between the number of useful bits n and the codeword length say $n_{\pi(j)}$ for the $\pi(j)^{th}$ channel. This code maps the sequence $\mathbf{x}_{(k)}^n$ to an incremental codeword having $\mathbf{x}_{(k)}^n$ as systematic symbols and $\mathbf{p}_{(k)}$ as parity symbols. The parity vectors are sequential: $\mathbf{p}_{(k)} = (\mathbf{p}_{(k)}^{nr_1}, \mathbf{p}_{(k)}^{nr_2}, \dots, \mathbf{p}_{(k)}^{nr_{M(k)}})$, where the sum rate is given by $\sum_{i=1}^{M(k)} r_i \geq S_{(k)}$ defined in (9). The compression rates r_m and channel codes R_m^{ch} are related through

$$\sum_{i=1}^{\pi(j)} r_i = \frac{n_{\pi(j)} - n}{n} = \frac{1}{R_{\pi(j)}^{ch}} - 1, \forall j \in \mathcal{J}_{(k)}. \quad (87)$$

The whole parity vector is stored with markers to identify the end of each subvector $\mathbf{p}_{(k)}^{nr_{\pi(j)}}$. Then, upon request of source (k) while the previous request was (j^*) , the transmitter sends the vector $(\mathbf{p}_{(k)}^{nr_1}, \mathbf{p}_{(k)}^{nr_2}, \dots, \mathbf{p}_{(k)}^{nr_{\pi(j^*)}})$. Finally, the storage and transmission cost can be related to the channel rates:

$$\text{practical transmission rate} = \sum_{i=1}^{\pi(j^*)} r_i = \frac{1}{R_{\pi(j^*)}^{ch}} - 1 \quad (88)$$

$$\text{practical storage rate} = \sum_{i=1}^{M(k)} r_i = \frac{1}{R_{M(k)}^{ch}} - 1 \quad (89)$$

Following the proof in [14], one can show that if the channel code achieves the capacity of the equivalent channel, then the practical scheme will also achieve the bounds derived in Theorem 5. In practice, the sequential parity vector can be obtained from LDPC code constructions such as LDPC staircase codes and LDPCA codes, as we illustrate in the next section.

The coding strategy relies on the fact that the probability of decoding error converges to 0 when $n \rightarrow \infty$. In other words, the probability of error may not always be 0 when encoding a finite length source, which could be damageable when using, in turn, this source as side information for the decoding of another one (consecutively requested by the same client). To circumvent this issue in practice, we propose to simulate the decoding at the encoding stage and to add extra parity bits until the probability of error actually reaches 0. This results in extra storage and transmission rates but prevents any error propagation during the client's navigation. However, due to the optimal performances obtained in the theoretical study (see Theorem 5), we know that this extra rate converges to 0 when $n \rightarrow \infty$.

	Number of symbols sent						Number Stored
	$p_{j_1} = 0.01$	$p_{j_2} = 0.05$	$p_{j_3} = 0.1$	$p_{j_4} = 0.15$	$p_{j_5} = 0.2$	$p_{j_6} = 0.25$	
theoretical	400	2,000	4,000	6,000	8,000	10,000	40,000
LDPC-Staircase	423	2,023	4,023	6,023	8,023	10,023	40,023

$n = 40000$

TABLE II: Comparison between theoretical and practical rate-storage performance for the coding of a source $X_{(k)}$ made of n symbols. The set of possible side information sources $\mathcal{J}_{(k)}$ is made of 6 sources, with different erasure probability. The results were obtained with a LDPC staircase coder [3].

	Number of symbols sent						Number Stored
	$p_{j_1} = 0.01$	$p_{j_2} = 0.05$	$p_{j_3} = 0.1$	$p_{j_4} = 0.15$	$p_{j_5} = 0.2$	$p_{j_6} = 0.25$	
theoretical	32	114	186	242	286	322	322
LDPCA	52	147	217	283	332	377	377

(a) $n = 396$

	Number of symbols sent						Number Stored
	$p_{j_1} = 0.01$	$p_{j_2} = 0.05$	$p_{j_3} = 0.1$	$p_{j_4} = 0.15$	$p_{j_5} = 0.2$	$p_{j_6} = 0.25$	
theoretical	512	1815	2972	3864	4575	5141	5141
LDPCA	913	2160	3325	4264	5194	6329	6329

(b) $n = 6336$

TABLE III: Comparison between theoretical and practical rate-storage performance for the coding of a source $X_{(k)}$ made of n symbols. The set of possible side information sources $\mathcal{J}_{(k)}$ is made of 6 sources, with different error probability. The results were obtained with a LDPCA coder [4].

IX. EXPERIMENTAL VALIDATION

A. Experimental setup

The rate-storage bounds derived in this paper state the performance that a SMRA scheme can theoretically achieve. Building SMRA codes for general sources is another complex research problem and we leave it for future works. However, in this section, we show that for binary i.i.d. uniform sources, already existing codes, namely LDPC staircase [3] and LDPCA [4], can help building efficient practical implementation of SMRA schemes. For this experimental validation, we thus assume that the $X_{(\ell)}$ are binary i.i.d. uniform sources. As in the theoretical results derived in previous section, let us focus on the performance of a single source $X_{(k)}$ that is assumed to be requested by the user. For binary i.i.d. sources, we consider two forms for the conditional probability distribution $P(x_{(k)}|x_{(j)})$ (with $j \in \mathcal{J}_{(k)}$) that correspond to either erasures or binary errors. In the next two sections, we study each of the two distributions and propose a practical SMRA scheme for both cases.

B. Erasure model

In an erasure channel with input $X_{(k)}$ and output $X_{(j)}$, the n -th component of $X_{(j)}$ is either equal to the n -th component of $X_{(k)}$, or erased with probability p . The rate needed to decode $X_{(k)}$ from $X_{(j)}$ depends on the probability p that an element is erased. We consider that the set $\mathcal{J}_{(k)}$ is made of 6 sources $X_{(j_1)}, \dots, X_{(j_6)}$, with respective erasure probabilities $p_{(j_1)} = 0.01$, $p_{(j_2)} = 0.05$, $p_{(j_3)} = 0.1$, $p_{(j_4)} = 0.15$, $p_{(j_5)} = 0.2$ and $p_{(j_6)} = 0.25$. In our experiment, we consider $n = 40000$. The source $X_{(k)}$ is encoded with an LDPC staircase code [3]. The systematic bits are discarded and only the parity bits are stored on the server. At the decoder, the parity bits are used to build $X_{(k)}$ using the source $X_{(j^*)}$ in memory as side information, *i.e.*, taking it for systematic bits and using the received parity bits to complete the codeword. Regarding the stored bitstream, the least correlated source $X_{(j_6)}$ corresponds to the highest bitrate and we store the amount of bits that leads to a very low error reconstruction of $X_{(k)}$ from $X_{(j_6)}$. At the transmission stage, for each source $X_{(j)}$ in memory, we extract from this bitstream the minimum amount of bits that are required to obtain $X_{(k)}$ losslessly. The obtained rates are shown in Table II. We see that the bitrate obtained with the LDPC staircase are close to the theoretical bounds. An important observation is also that from the same stored bitstream, a user is able to extract the piece of information that is required by its own navigation. In other words, the storage considers the worst case, but the rate is actually not penalized by the randomness of the navigation.

C. Binary error model

Consider a binary symmetric channel with input $X_{(k)}$, output $X_{(j)}$, and error probability $p = \mathbb{P}(X_{(k)} \neq X_{(j)})$. As before, we consider that the set $\mathcal{J}_{(k)}$ is made of 6 sources $X_{(j_1)}, \dots, X_{(j_6)}$, with respective error probabilities $p_{(j_1)} = 0.01$, $p_{(j_2)} = 0.05$, $p_{(j_3)} = 0.1$, $p_{(j_4)} = 0.15$, $p_{(j_5)} = 0.2$ and $p_{(j_6)} = 0.25$. The compression rate depends on the error probability through the formula $H = -p \log_2 p - (1-p) \log_2 (1-p)$. For this channel, we encode the source $X_{(k)}$ with an LDPCA code [4]. As before, the parity bits of the least correlated source $X_{(j_6)}$ are stored and contain the parity bits of all the sources. From this bitstream, a part of the parity bits are extracted to transmit $X_{(k)}$ given that one of the sources $X_{(j_1)}, \dots, X_{(j_6)}$ is available as side information at the decoder. The obtained rate are shown in Table III for the respective number of symbols $n = 396$ and $n = 6336$.

As in the case of the erasure channel, the results of Table III show that here again, the proposed implementation enables to extract a sub-bitstream that depends on the source $X_{(j)}$ and to reach a very low reconstruction error probability of $X_{(k)}$. This important result validates the idea that the rate is not penalized by the randomness of the user navigation.

X. CONCLUSION

In this paper, we have introduced and formalized the Sequential Massive Random Access problem, that deals with a set of sources encoded once for all on a server, followed by a partial transmission of the requested sources only. We have introduced a coding scheme in two parts that are the offline encoder for source storage, and the online extractor for transmission of the requested sources only. From both theoretical and simulation results, we have shown that, for a reasonable storage cost, the partial transmission of the requested sources can be done with the same rate as if re-encoding was allowed during the online phase. This fundamental result may be transposed to all the practical SMRA scenarios, by deriving specific correlation models coupled with efficient channel codes. Some other interesting aspects of SMRA might be further developed too, such as the influence of the navigation graph on the storage and transmission rates.

ACKNOWLEDGEMENT

This work has received a French government support granted to the Cominlabs excellence laboratory and managed by the National Research Agency in the “Investing for the Future” program under reference ANR-10-LABX-07-01.

REFERENCES

- [1] M. Hilbert and P. Lopez, “The worlds technological capacity to store, communicate, and compute information,” *Science*, vol. 332, no. 6025, pp. 60–65, Apr. 2011.
- [2] M. Tanimoto, “FTV: Free-viewpoint television,” *IEEE Signal Processing Magazine*, vol. 27, no. 6, pp. 555–570, Jul. 2012.
- [3] V. Roca, M. Cunche, and J. Lacan, *LDPC-Staircase Forward Error Correction (FEC) Schemes for FECFRAME*, IETF Std. RFC 6816, Dec. 2012. [Online]. Available: <http://datatracker.ietf.org/doc/rfc6816/>
- [4] D. Varodayan, A. Aaron, and B. Girod, “Rate-adaptive codes for distributed source coding,” *EURASIP Signal Processing*, vol. 86, no. 11, pp. 3123–3130, 2006.
- [5] M. A. Maddah-Ali and U. Niesen, “Fundamental limits of caching,” *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2856–2867, 2014.
- [6] P. Hassanzadeh, A. Tulino, J. Llorca, and E. Erkip, “Cache-aided coded multicast for correlated sources,” *arXiv preprint arXiv:1609.05831*, 2016.
- [7] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, “Network coding for distributed storage systems,” *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4539–4551, 2010.
- [8] A. El Gamal and T. Cover, “Achievable rates for multiple descriptions,” *IEEE transaction on Information Theory*, vol. 28, no. 6, pp. 851–857, Nov 1982.
- [9] S. N. Diggavi and V. A. Vaishampayan, “On multiple description source coding with decoder side information,” in *IEEE Information Theory Workshop*, 2004, pp. 88–93.
- [10] A. Sgarro, “Source coding with side information at several decoders,” *IEEE Transactions on Information Theory*, vol. 23, no. 2, pp. 179–182, 1977.
- [11] S. C. Draper, “Universal incremental slepian-wolf coding,” in *Allerton Conference on Communication, control and computing*, 2004, pp. 1757 – 1761.
- [12] T. Maugey and P. Frossard, “Interactive multiview video system with low complexity 2d look around at decoder,” *IEEE Transactions on Multimedia*, vol. 15, pp. 1–13, Aug. 2013.
- [13] D. Slepian and J. Wolf, “Noiseless coding of correlated information sources,” *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, July 1973.
- [14] A. Wyner, “Recent Result in the Shannon theory,” *IEEE Transactions on Information Theory*, vol. 20, no. 1, pp. 2–10, 1974.
- [15] S. C. Draper and E. Martinian, “Compound conditional source coding, Slepian-Wolf list decoding, and applications to media coding,” in *IEEE International Symposium on Information Theory*, 2007.
- [16] G. Cheung, A. Ortega, and N. Cheung, “Interactive streaming of stored multiview video using redundant frame structures,” *IEEE Transactions on Image Processing*, vol. 3, no. 3, pp. 744–761, Mar. 2011.
- [17] E. Yang and D. He, “Interactive encoding and decoding for one way learning: Near lossless recovery with side information at the decoder,” *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1808–1824, 2010.
- [18] T. Cover and J. Thomas, *Elements of information theory, second Edition*. Wiley, 2006.
- [19] R. W. Yeung, *Information Theory and Network Coding*. Springer, 2010.
- [20] T. Han, *Information-spectrum methods in information theory*. Springer, 2003.
- [21] P. Moulin and Y. Wang, “Capacity and random-coding exponents for channel coding with side information,” *IEEE Transactions on Information Theory*, vol. 53, no. 5, pp. 1326–1347, 2007.
- [22] E. Dupraz, A. Roumy, and M. Kieffer, “Source coding with side information at the decoder and uncertain knowledge of the correlation,” *IEEE Transactions on Communications*, vol. 62, no. 1, pp. 269 – 279, Jan. 2014.
- [23] C. Tian and S. N. Diggavi, “On multistage successive refinement for wyner-ziv source coding with degraded side informations,” *IEEE Transactions on Information Theory*, vol. 53, no. 8, pp. 2946–2960, 2007.
- [24] S. Watanabe and S. Kuzuoka, “Universal Wyner-Ziv coding for distortion constrained general side information,” *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 7568–7583, 2014.
- [25] M. Feder and N. Shulman, “Source broadcasting with unknown amount of receiver side information,” in *Information Theory Workshop (ITW), Proceedings.*, march 2002, pp. 302 – 311.
- [26] N. Shulman, “Communication over an unknown channel via common broadcasting,” Ph.D. dissertation, Tel Aviv University, 2003.
- [27] D. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.
- [28] —, “Fountain Codes,” *IEE Proceedings on Communications*, vol. 152, no. 6, pp. 1062–1068, Dec. 2005.
- [29] P. L. Dragotti and M. Gastpar, *Distributed Source Coding: Theory, Algorithms and Applications*. Academic Press, 2009.